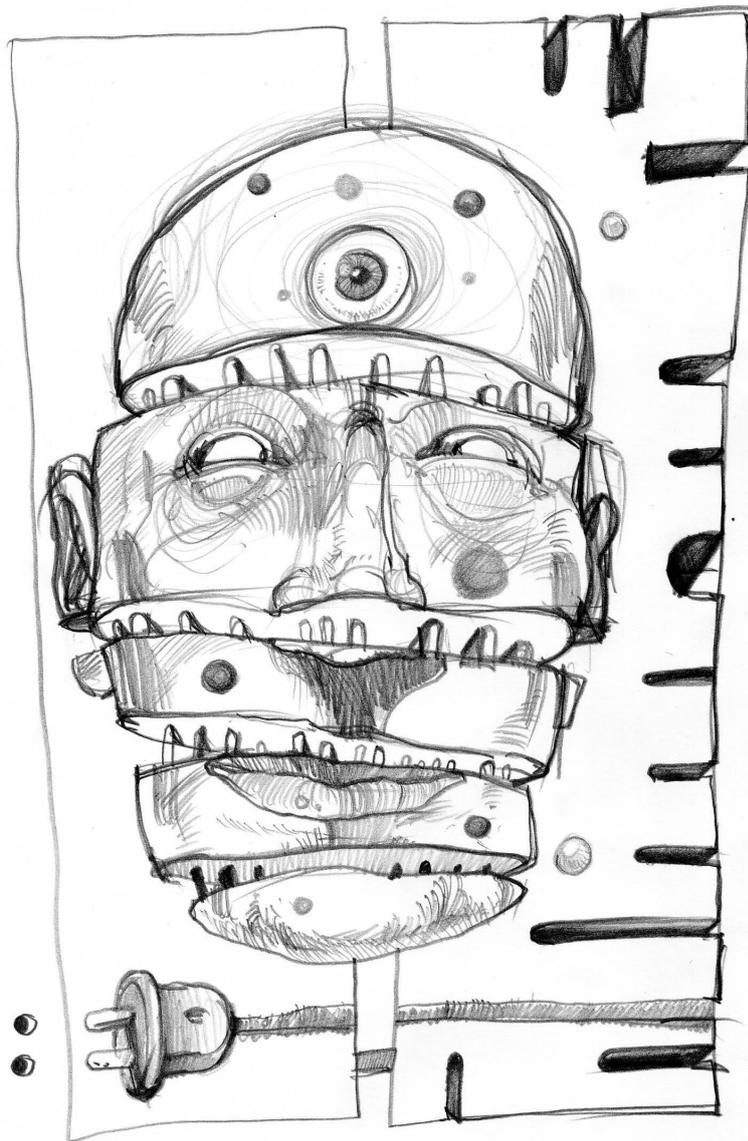


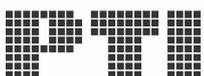
Annals of Computer Science and Information Systems  
Volume 18

# Proceedings of the 2019 Federated Conference on Computer Science and Information Systems

September 1–4, 2019. Leipzig, Germany



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





# Annals of Computer Science and Information Systems, Volume 18

## Series editors:

Maria Ganzha (Editor-in-Chief),

*Systems Research Institute Polish Academy of Sciences and Warsaw University of Technology, Poland*

Leszek Maciaszek,

*Wrocław University of Economy, Poland and Macquarie University, Australia*

Marcin Paprzycki,

*Systems Research Institute Polish Academy of Sciences and Management Academy, Poland*

## Senior Editorial Board:

Wil van der Aalst,

*Department of Mathematics & Computer Science, Technische Universiteit Eindhoven (TU/e), Eindhoven, Netherlands*

Marco Aiello,

*Faculty of Mathematics and Natural Sciences, Distributed Systems, University of Groningen, Groningen, Netherlands*

Mohammed Atiquzzaman,

*School of Computer Science, University of Oklahoma, Norman, USA*

Jan Bosch,

*Chalmers University of Technology, Gothenburg, Sweden*

Barrett Bryant,

*Department of Computer Science and Engineering, University of North Texas, Denton, USA*

Włodzisław Duch,

*Department of Informatics, and NeuroCognitive Laboratory, Center for Modern Interdisciplinary Technologies, Nicolaus Copernicus University, Toruń, Poland*

Ana Fred,

*Department of Electrical and Computer Engineering, Instituto Superior Técnico (IST—Technical University of Lisbon), Lisbon, Portugal*

Janusz Górski,

*Department of Software Engineering, Gdańsk University of Technology, Gdańsk, Poland*

Giancarlo Guizzardi,

*Free University of Bolzano-Bozen, Italy, Senior Member of the Ontology and Conceptual Modeling Research Group (NEMO), Brazil*

Francisco Herrera,

*Dept. Computer Sciences and Artificial Intelligence Andalusian Research Institute in Data Science and Computational Intelligence (DaSCI) University of Granada, Spain*

Mike Hinchey,

*Lero—the Irish Software Engineering Research Centre, University of Limerick, Ireland*

Janusz Kacprzyk,

*Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland*

Irwin King,

*The Chinese University of Hong Kong, Hong Kong*

Juliusz L. Kulikowski,

*Nalęcz Institute of Biocybernetics and Biomedical Engineering, Polish Academy of Sciences, Warsaw, Poland*

Michael Luck,

*Department of Informatics, King's College London, London, United Kingdom*

Jan Madey,

*Faculty of Mathematics, Informatics and Mechanics at the University of Warsaw, Poland*

Stan Matwin,

*Dalhousie University, University of Ottawa, Canada and Institute of Computer Science,  
Polish Academy of Science, Poland*

Marjan Mernik,

*University of Maribor, Slovenia*

Michael Segal,

*Ben-Gurion University of the Negev, Israel*

Andrzej Skowron,

*Faculty of Mathematics, Informatics and Mechanics at the University of Warsaw, Poland*

John F. Sowa,

*VivoMind Research, LLC, USA*

George Spanoudakis,

*Research Centre for Adaptive Computing Systems (CeNACS), School of Mathematics,  
Computer Science and Engineering, City, University of London*

**Editorial Associates:**

Katarzyna Wasielewska,

*Systems Research Institute Polish Academy of Sciences, Poland*

Paweł Sitek,

*Kielce University of Technology, Kielce, Poland*

**T<sub>E</sub>Xnical editor:** Aleksander Denisiuk,

*University of Warmia and Mazury in Olsztyn, Poland*

# Proceedings of the 2019 Federated Conference on Computer Science and Information Systems

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



2019, Warszawa,  
Polskie Towarzystwo  
Informatyczne



2019, New York City,  
Institute of Electrical and  
Electronics Engineers

Annals of Computer Science and Information Systems, Volume 18  
Proceedings of the 2019 Federated Conference on Computer Science and  
Information Systems

ART: ISBN 978-83-955416-0-5, IEEE Catalog Number CFP1985N-ART

USB: ISBN 978-83-952357-9-5, IEEE Catalog Number CFP1985N-USB

WEB: ISBN 978-83-952357-8-8

ISSN 2300-5963

DOI 10.15439/978-83-952357-8-8

© 2019, Polskie Towarzystwo Informatyczne

Ul. Solec 38/103

00-394 Warsaw, Poland

**Contact:** [secretariat@fedcsis.org](mailto:secretariat@fedcsis.org)

<http://annals-csis.org/>

**Cover art:**

Janusz Kozak,

*Elbląg, Poland*

**Also in this series:**

Volume 20: Communication Papers of the 2019 Federated Conference on Computer  
Science and Information Systems, **ISBN WEB: 978-83-955416-3-6, ISBN USB: 978-83-955416-4-3**

Volume 19: Position Papers of the 2019 Federated Conference on Computer Science and  
Information Systems, **ISBN WEB: 978-83-955416-1-2, ISBN USB: 978-83-955416-2-9**

Volume 17: Communication Papers of the 2018 Federated Conference on Computer  
Science and Information Systems, **ISBN WEB: 978-83-952357-0-2, ISBN USB: 978-83-952357-1-9**

Volume 16: Position Papers of the 2018 Federated Conference on Computer Science and  
Information Systems, **ISBN WEB: 978-83-949419-8-7, ISBN USB: 978-83-949419-9-4**

Volume 15: Proceedings of the 2018 Federated Conference on Computer Science and  
Information Systems, **ISBN WEB: 978-83-949419-5-6, ISBN USB: 978-83-949419-6-3,**

**ISBN ART: 978-83-949419-7-0**

Volume 14: Proceedings of the First International Conference on Information  
Technology and Knowledge Management, **ISBN WEB: 978-83-949419-2-5,**

**ISBN USB: 978-83-949419-1-8, ISBN ART: 978-83-949419-0-1**

Volume 13: Communication Papers of the 2017 Federated Conference on Computer  
Science and Information Systems, **ISBN WEB: 978-83-922646-2-0, ISBN USB: 978-83-922646-3-7**

Volume 12: Position Papers of the 2017 Federated Conference on Computer Science and  
Information Systems, **ISBN WEB: 978-83-922646-0-6, ISBN USB: 978-83-922646-1-3**

Volume 11: Proceedings of the 2017 Federated Conference on Computer Science and  
Information Systems, **ISBN WEB: 978-83-946253-7-5, ISBN USB: 978-83-946253-8-2,**

**ISBN ART: 978-83-946253-9-9**

Volume 10: Proceedings of the Second International Conference on Research in  
Intelligent and Computing in Engineering, **ISBN WEB: 978-83-65750-05-1,**

**ISBN USB: 978-83-65750-06-8**

Volume 9: Position Papers of the 2016 Federated Conference on Computer Science and  
Information Systems, **ISBN WEB: 978-83-60810-93-4, ISBN USB: 978-83-60810-94-1**

Volume 8: Proceedings of the 2016 Federated Conference on Computer Science and  
Information Systems, **ISBN WEB: 978-83-60810-90-3, ISBN USB: 978-83-60810-91-0,**

**ISBN ART: 978-83-60910-92-7**

DEAR Reader, it is our pleasure to present to you Proceedings of the 2019 Federated Conference on Computer Science and Information Systems (FedCSIS), which took place in Leipzig, Germany, on September 1-4, 2019.

FedCSIS 2019 was Chaired by prof. Bogdan Franczyk, while prof. Rainer Unland acted as the Chair of the Organizing Committee. This year, FedCSIS was organized by the Polish Information Processing Society (Mazovia Chapter), IEEE Poland Section Computer Society Chapter, Systems Research Institute Polish Academy of Sciences, Warsaw University of Technology, Wrocław University of Economics, and Leipzig University, Germany.

FedCSIS 2019 was technically co-sponsored by: IEEE Region 8, IEEE Poland Section, IEEE Computer Society Technical Committee on Intelligent Informatics, IEEE Czechoslovakia Section Computer Society Chapter, IEEE Poland Section Gdańsk Computer Society Chapter, IEEE Poland Section Systems, Man, and Cybernetics Society Chapter, IEEE Poland Section Control System Society Chapter, IEEE Poland Section Computational Intelligence Society Chapter, Committee of Computer Science of the Polish Academy of Sciences, Polish Operational and Systems Research Society, Mazovia Cluster ICT Poland and Eastern Cluster ICT Poland. FedCSIS 2019 was sponsored by Intel.

During FedCSIS 2019, keynote lectures were delivered by:

- Enrique Alba, University of Málaga, Spain, “*Intelligent Systems for Smart Cities*”
- Francisco Herrera, Dept. Computer Sciences and Artificial Intelligence Andalusian Research Institute in Data Science and Computational Intelligence (DaSCI) University of Granada, “*Deep Data and Big Learning: More quality data for better knowledge*”
- George Spanoudakis, Research Centre for Adaptive Computing Systems (CeNACS), School of Mathematics, Computer Science and Engineering, City, University of London, “*Cyber security risks: Comprehensive mitigation through technical, contractual and financial mitigation mechanisms*”

FedCSIS 2019 consisted of five Tracks and a Doctoral Symposium. Tracks were divided into Technical Sessions. Sessions preannounced in Call for Papers as track-related events (conferences, symposia, workshops, special sessions).

- **Track 1: Artificial Intelligence and Applications**
  - Advances in Artificial Intelligence and Applications (14th Symposium AAIA'19)
  - Computational Optimization (12th Workshop WCO'19)
  - Smart Energy Networks & Multi-Agent Systems (7th Workshop SEN-MAS'19)
- **Track 2: Computer Science & Systems**
  - Computer Aspects of Numerical Algorithms (12th Workshop CANA'19)
  - Cryptography and Security Systems (6th Conference C&SS'19)
  - Language Technologies and Applications (4th Workshop LTA'19)
  - Multimedia Applications and Processing

(12th Symposium MMAP'19)

- Advances in Programming Languages (7<sup>th</sup> Workshop WAPL'19)
- Scalable Computing (10th Workshop WSC'19)
- **Track 3: Network Systems and Applications**
  - Advances in Network Systems and Applications (ANSA)
  - Internet of Things - Enablers, Challenges and Applications (3rd Workshop IoT-ECAW'19)
- **Track 4: Information Systems and Technology**
  - Advanced Information Technologies for Management (16th Conference AITM'19)
  - Data Science in Health (1st Special Session DSH'19)
  - Data Analysis and Computation for Digital Ecosystems (1st Workshop InC2Eco'19)
  - Information Systems Management (14th Conference ISM'19)
  - Knowledge Acquisition and Management (25th Conference KAM'19)
- **Track 5: Software and System Engineering**
  - Advances in Software and System Engineering (ASSE)
  - Cyber-Physical Systems (6th Workshop IWCPs-6)
  - Lean and Agile Software Development (3rd International Conference LASD'19)
  - Multimedia, Interaction, Design and Innovation (7th Conference MIDI'19)
  - Software Engineering (39th IEEE Workshop SEW-39)
- **DS-RAIT'19 - 6th Doctoral Symposium on Recent Advances in Information Technology**

The 2019 edition of an AAIA'19 Data Mining Challenge was called *Clash Royale Challenge: How to Select Training Decks for Win-rate*. ~~This year the~~ The task was related to the problem of selecting an optimal training data subset for learning how to predict win-rates of the most popular *Clash Royale decks*. Awards for the winners of the contest were sponsored by: Esensei and the Mazovia Chapter of the Polish Information Processing Society. Papers resulting from the competition are included in the Conference Proceedings (Chapter of Track 1: AAIA).

Each paper, found in this volume, was refereed by at least two referees and the acceptance rate of regular full papers was ~20,8% (62 regular full papers out of 298 general submissions).

The program of FedCSIS required a dedicated effort of many people. Each event constituting FedCSIS had its own Organizing and Program Committee. We would like to express our warmest gratitude to all Committee members for their hard work in attracting and later refereeing 302 submissions (regular and data mining).

We thank the authors of papers for their great contribution to research and practice in computing and information systems. We thank the invited speakers for sharing their knowledge and wisdom with the participants. Finally, we thank all those responsible for staging the conference in Leipzig. Or-

ganizing a conference of this scope and level could only be achieved by the collaborative effort of a highly capable team taking charge of such matters as conference registration system, finances, the venue, social events, catering, handling all sorts of individual requests from the authors, preparing the conference rooms, etc.

We hope you had an inspiring conference and an unforgettable stay in the beautiful city of Leipzig. We also hope to meet you again for FedCSIS 2020 in Sofia, Bulgaria.

**Co-Chairs of the FedCSIS Conference Series**

***Maria Ganzha***, *Warsaw University of Technology, Poland and Systems Research Institute Polish Academy of Sciences, Warsaw, Poland*

***Leszek Maciaszek***, *Wroclaw University of Economics, Wroclaw, Poland and Macquarie University, Sydney, Australia*

***Marcin Paprzycki***, *Systems Research Institute Polish Academy of Sciences, Warsaw Poland and Management Academy, Warsaw, Poland*

# Proceedings of the Federated Conference on Computer Science and Information Systems

September 1–4, 2019. Leipzig, Germany

---

## TABLE OF CONTENTS

---

---

### ARTIFICIAL INTELLIGENCE AND APPLICATIONS

---

---

#### 14<sup>TH</sup> INTERNATIONAL SYMPOSIUM ADVANCES IN ARTIFICIAL INTELLIGENCE AND APPLICATIONS

---

<b>Call For Papers</b>	<b>1</b>
<b>Clash Royale Challenge: How to Select Training Decks for Win-rate Prediction</b> <i>Andrzej Janusz, Łukasz Grad, Marek Grzegorowski</i>	<b>3</b>
<b>Greedy Incremental Support Vector Regression</b> <i>Dymitr Ruta, Ling Cen, Quang Hieu Vu</i>	<b>7</b>
<b>Training Subset Selection for Support Vector Regression</b> <i>Cenru Liu, Jiahao Cen</i>	<b>11</b>
<b>Efficient Support Vector Regression with Reduced Training Data</b> <i>Ling Cen, Quang Hieu Vu, Dymitr Ruta</i>	<b>15</b>
<b>Information granule system induced by a perceptual system</b> <i>Anna Bryniarska</i>	<b>19</b>
<b>Improving Real-Time Performance of U-Nets for Machine Vision in Laser Process Control</b> <i>Przemysław Dolata, Jacek Reiner</i>	<b>29</b>
<b>Predicting blood glucose using an LSTM Neural Network</b> <i>Touria El Idrissi, Ali Idri, Ibtissam Abnane, Zohra Bakkoury</i>	<b>35</b>
<b>Accurate Retrieval of Corporate Reputation from Online Media Using Machine Learning</b> <i>Achim Klein, Martin Riekert, Velizar Dinev</i>	<b>43</b>
<b>A Specialized Evolutionary Approach to the bi-objective Travelling Thief Problem</b> <i>Maciej Laszczyk, Paweł B. Myszkowski</i>	<b>47</b>
<b>Urban Sound Classification using Long Short-Term Memory Neural Network</b> <i>Iurii Lezhenin, Natalia Bogach, Evgeny Pyshkin</i>	<b>57</b>
<b>Quantitative Impact of Label Noise on the Quality of Segmentation of Brain Tumors on MRI scans</b> <i>Michał Marcinkiewicz, Grzegorz Mrukwa</i>	<b>61</b>
<b>Non-dominated Sorting Tournament Genetic Algorithm for Multi-Objective Travelling Salesman Problem</b> <i>Paweł B. Myszkowski, Maciej Laszczyk, Kamil Dziadek</i>	<b>67</b>
<b>Development of a Flexible Mizar Tokenizer and Parser for Information Retrieval System</b> <i>Kazuhisa Nakasho</i>	<b>77</b>

<b>British Sign Language Recognition In The Wild Based On Multi-Class SVM</b>	<b>81</b>
<i>Joanna Isabelle Olszewska, M. Quinn</i>	
<b>Counting Instances of Objects Specified By Vague Locations Using Neural Networks on Example of Honey Bees</b>	<b>87</b>
<i>Jerzy Respondek, Weronika Westwańska</i>	
<b>Generating Human Mobility Route Based on Generative Adversarial Network</b>	<b>91</b>
<i>Ha Yoon Song, Moo Sang Baek, Minsuk Sung</i>	
<b>A Deep Learning and Multimodal Ambient Sensing Framework for Human Activity Recognition</b>	<b>101</b>
<i>Ali Yachir, Abdenour Amamra, Badis Djamaa, Ali Zerrouki, Ahmed khierEddine Amour</i>	
<b>Machine-learning at the service of plastic surgery: a case study evaluating facial attractiveness and emotions using R language</b>	<b>107</b>
<i>Lubomír Štěpánek, Pavel Kasal, Jan Měšťák</i>	

---

## 12<sup>TH</sup> INTERNATIONAL WORKSHOP ON COMPUTATIONAL OPTIMIZATION

---

<b>Call For Papers</b>	<b>113</b>
<b>A Minimum Set-Cover Problem with several constraints</b>	<b>115</b>
<i>Jens Dörpinghaus, Carsten Düing, Vera Weil</i>	
<b>KPIs for Optimal Location of charging stations for Electric Vehicles: the Biella case-study</b>	<b>123</b>
<i>Edoardo Fadda, Daniele Manerba, Roberto Tadei, Paolo Camurati, Gianpiero Cabodi</i>	
<b>A novel integer linear programming model for routing and spectrum assignment in optical networks</b>	<b>127</b>
<i>Youssef Hadhbi, Hervé Kerivin, Annegret Wagler</i>	
<b>An Efficient Exhaustive Search for the Discretizable Distance Geometry Problem with Interval Data</b>	<b>135</b>
<i>Antonio Mucherino, Jung-Hsin Lin</i>	
<b>Models and Algorithms for Natural Disaster Evacuation Problems</b>	<b>143</b>
<i>Alain Quiliot, Christian Artigues, Emmanuel Hebrard, H�el�ene Toussaint</i>	
<b>Best Response Dynamics for VLSI Physical Design Placement</b>	<b>147</b>
<i>Michael Rapoport, Tami Tamir</i>	
<b>Integration of Polynomials over n-Dimensional Simplices</b>	<b>157</b>
<i>Abdenebi Rouigueb, Mohamed Maiza, Abderahmane Tkourt, Imed Cherchour</i>	
<b>Customized Genetic Algorithm for Facility Allocation using p-median</b>	<b>165</b>
<i>Sergio Silva, Marly Costa, Cicero Costa Filho</i>	
<b>An algorithm for 1-space bounded cube packing</b>	<b>171</b>
<i>Lukasz Zielonka</i>	
<b>Ant Colony Optimization Algorithm for Workforce Planning: Influence of the Evaporation Parameter</b>	<b>177</b>
<i>Stefka Fidanova, Gabriel Luque, Olympia Roeva, Maria Ganzha</i>	

---

## 7<sup>TH</sup> INTERNATIONAL WORKSHOP ON SMART ENERGY NETWORKS & MULTI-AGENT SYSTEMS

---

<b>Call For Papers</b>	<b>183</b>
<b>Tool-assisted Surrogate Selection for Simulation Models in Energy Systems</b>	<b>185</b>
<i>Stephan Balduin, Frauke Oest, Marita Blank-Babazadeh, Astrid Nie�e, Sebastian Lehnhoff</i>	
<b>Towards fully Decentralized Multi-Objective Energy Scheduling</b>	<b>193</b>
<i>Joerg Bremer, Sebastian Lehnhoff</i>	

---

## COMPUTER SCIENCE & SYSTEMS

---

<b>Call For Papers</b>	<b>203</b>
------------------------	------------

---

## 12<sup>TH</sup> WORKSHOP ON COMPUTER ASPECTS OF NUMERICAL ALGORITHMS

---

<b>Call For Papers</b>	<b>205</b>
<b>Parallel cache-efficient code for computing the McCaskill partition functions</b> <i>Marek Pałkowski, Włodzimierz Bielecki</i>	<b>207</b>

---

## 6<sup>TH</sup> INTERNATIONAL CONFERENCE ON CRYPTOGRAPHY AND SECURITY SYSTEMS

---

<b>Call For Papers</b>	<b>211</b>
<b>The Low-Area FPGA Design for the Post-Quantum Cryptography Proposal Round5</b> <i>Michał Andrzejczak</i>	<b>213</b>
<b>Accelerating Multivariate Cryptography with Constructive Affine Stream Transformations</b> <i>Michael Carenzo, Monika Polak</i>	<b>221</b>
<b>Spline-Wavelet Bent Robust Codes</b> <i>Alla Levina, Gleb Ryaskin, Igor Zikratov</i>	<b>227</b>
<b>Cryptographic keys management system based on DNA strands</b> <i>Marek Miśkiewicz, Bogdan Książopolski</i>	<b>231</b>
<b>Malicious and Harmless Software in the Domain of System Utilities</b> <i>Jana Št'astná</i>	<b>237</b>

---

## 4<sup>TH</sup> INTERNATIONAL WORKSHOP ON LANGUAGE TECHNOLOGIES AND APPLICATIONS

---

<b>Call For Papers</b>	<b>247</b>
<b>Multilingual Knowledge Base Completion by Cross-lingual Semantic Relation Inference</b> <i>Nadia Bebeshina-Clairet, Mathieu Lafourcade</i>	<b>249</b>
<b>Deep Learning Hyper-parameter Tuning for Sentiment Analysis in Twitter based on Evolutionary Algorithms</b> <i>Eugenio Martínez Cámara, Nuria Rodríguez Barroso, Antonio R. Moya, José Alberto Fernández, Elena Romero, Francisco Herrera</i>	<b>255</b>
<b>Knowledge Extraction and Applications utilizing Context Data in Knowledge Graphs</b> <i>Jens Dörpinghaus, Andreas Stefan</i>	<b>265</b>
<b>Towards semantic-rich word embeddings</b> <i>Grzegorz Beringer, Mateusz Jabłoński, Piotr Januszewski, Andrzej Sobecki, Julian Szymański</i>	<b>273</b>
<b>Languages' Impact on Emotional Classification Methods</b> <i>Alexander Christoffer Eilertsen, Dennis Højbjerg Rose, Peter Langballe Erichsen, Rasmus Engesgaard Christensen, Rudra Pratap Deb Nath</i>	<b>277</b>
<b>Signature analysis system using a convolutional neural network</b> <i>Alicja Winnicka, Karolina Kęsik, Dawid Połap</i>	<b>287</b>

---

## 12<sup>TH</sup> INTERNATIONAL SYMPOSIUM ON MULTIMEDIA APPLICATIONS AND PROCESSING

---

<b>Call For Papers</b>	<b>291</b>
<b>Creating See-Around Scenes using Panorama Stitching</b> <i>Saja Alferidah, Nora Alkhaldi</i>	<b>293</b>
<b>A Social Bonds Integration Approach for Crowd Panic Simulation</b> <i>Imene Bouderbal, Abdenour Amamra</i>	<b>303</b>
<b>GNSS-based Sound Card Synchronization</b> <i>Alexander Carôt, Hasan Mahmood, Christian Hoene</i>	<b>309</b>

<b>Palmprint Recognition Based on Convolutional Neural Network-Alexnet</b>	<b>313</b>
<i>Weyiyong Gong, Xinman Zhang, Bohua Deng, Xuebin Xu</i>	
<b>A Contribution to Workplace Ergonomics Evaluation Using Multimedia Tools and Virtual Reality</b>	<b>317</b>
<i>Roman Leskovský, Erik Kučera, Oto Haffner, Jakub Matišák, Danica Rosinová, Erich Stark</i>	
<b>Information theoretical secure key sharing protocol for noiseless public constant parameter channels without cryptographic assumptions</b>	<b>327</b>
<i>Guillermo Morales-Luna, Valery Korzhik, Vladimir Starostin, Muaed Kabardov, Aleksandr Gerasimovich, Victor Yakovlev, Aleksey Zhuvikin</i>	
<b>License Plate Detection with Machine Learning Without Using Number Recognition</b>	<b>333</b>
<i>Kazuo Ohzeki, Max Geigis, Stefan Schneider</i>	
<b>Depth Map Improvements for Stereo-based Depth Cameras on Drones</b>	<b>341</b>
<i>Daniel Pohl, Sergey Dorodnicov, Markus Ahtelik</i>	
<b>Comparison of singing voice quality from the beginning of the phonation and in the stable phase in the case of choral voices</b>	<b>349</b>
<i>Edward Pótrolniczak, Michał Kramarczyk</i>	
<b>Automatic Assessment of Narrative Answers Using Information Retrieval Techniques</b>	<b>355</b>
<i>Liana Stanescu, Benjamin Savu</i>	
<b>Fuzzy Logic PID Control of a PMDCM Speed Connected to a 10-kW DC PV Array Microgrid - Case Study</b>	<b>359</b>
<i>Roxana-Elena Tudoroiu, Mohammed Zaheeruddin, Nicolae Tudoroiu, Dumitru Dan Burdescu</i>	
<b>Object detection in the police surveillance scenario</b>	<b>363</b>
<i>Artur Wilkowski, Włodzimierz Kasprzak, Maciej Stefańczyk</i>	
<b>Robust Image Forgery Detection Using Point Feature Analysis</b>	<b>373</b>
<i>Youssef William, Sherine Safwat, Mohammed Abdel-Megeed Salem</i>	
<b>Weighted Multimodal Biometric Recognition Algorithm Based on Histogram of Contourlet Oriented Gradient Feature Description</b>	<b>381</b>
<i>Xinman Zhang, Dongxu Cheng, Xuebin Xu</i>	
<b>A GIS Data Realistic Road Generation Approach for Traffic Simulation</b>	<b>385</b>
<i>Yacine Amara, Abdenour Amamra, Yasmine Daheur, Lamia Saichi</i>	
<b>Crime Scene Reconstruction with RGB-D Sensors</b>	<b>391</b>
<i>Abdenour Amamra, Yacine Amara, Khalid Boumaza, Aissa Benayad</i>	
<hr/>	
<b>7<sup>TH</sup> WORKSHOP ON ADVANCES IN PROGRAMMING LANGUAGES</b>	
<hr/>	
<b>Call For Papers</b>	<b>397</b>
<b>Composition of Languages Embedded in Scala</b>	<b>399</b>
<i>Syed Hossein Haeri, Paul Keir</i>	
<b>Supporting Source Code Annotations with Metadata-Aware Development Environment</b>	<b>411</b>
<i>Ján Juhár</i>	
<hr/>	
<b>10<sup>TH</sup> WORKSHOP ON SCALABLE COMPUTING</b>	
<hr/>	
<b>Call For Papers</b>	<b>421</b>
<b>Measure of Adequacy for the Supercomputer Job Management System Model</b>	<b>423</b>
<i>Anton Baranov, Pavel Telegin, Boris Shabanov, Dmitriy Lyakhovets</i>	
<b>Towards Big Data Solutions for Industrial Tomography Data Processing</b>	<b>427</b>
<i>Aleksandra Kowalska, Piotr Łuczak, Dawid Sielski, Tomasz Kowalski, Andrzej Romanowski, Dominik Sankowski</i>	
<b>Whose Fault is It? Correctly Attributing Outages in Cloud Services</b>	<b>433</b>
<i>Maurizio Naldi, Matteo Adriani</i>	

---

## NETWORK SYSTEMS AND APPLICATIONS

---

---

### ADVANCES IN NETWORK SYSTEMS AND APPLICATIONS

---

<b>Call For Papers</b>	<b>441</b>
<b>Formalization of Software Risk Assessment Results in Legal Metrology Based on ISO/IEC 18045 Vulnerability Analysis</b>	<b>443</b>
<i>Marko Esche, Felix Salwiczek, Federico Grasso Toro</i>	
<b>Fane: A Firewall Appliance For The Smart Home</b>	<b>449</b>
<i>Christoph Haar, Erik Buchmann</i>	
<b>Standardized container virtualization approach for collecting host intrusion detection data</b>	<b>459</b>
<i>Martin Max Röhling, Martin Grimmer, Dennis Kreußel, Jörn Hoffmann, Bogdan Franczyk</i>	

---

### 3<sup>RD</sup> WORKSHOP ON INTERNET OF THINGS - ENABLERS, CHALLENGES AND APPLICATIONS

---

<b>Call For Papers</b>	<b>465</b>
<b>Remote Programming and Reconfiguration System for Embedded Devices</b>	<b>467</b>
<i>Robert Brzoza-Woch, Tomasz Michalec, Maksymilian Wojczuk, Tomasz Szydło</i>	
<b>A Framework for Autonomous UAV Swarm Behavior Simulation</b>	<b>471</b>
<i>Piotr Cybulski</i>	
<b>Using Relay Nodes in Wireless Sensor Networks: A Review</b>	<b>479</b>
<i>Mustapha Reda Senouci, Mostefa Zafer, Mohamed Aissani</i>	
<b>Inference of driver behavior using correlated IoT data from the vehicle telemetry and the driver mobile phone</b>	<b>487</b>
<i>Daniel Alves da Silva, José Alberto Sousa Torres, Alexandre Pinheiro, Francisco L. de Caldas Filho, Fabio L. L. Mendonça, Bruno J. G. Praciano, Guilherme Oliveira Kfourir, Rafael T. de Sousa Jr</i>	
<b>Smart Urban Design Space</b>	<b>493</b>
<i>Philipp Skowron, Michael Aleithe, Susanne Wallrafen, Marvin Hubl, Julian Fietkau, Bogdan Franczyk</i>	
<b>An Adaptation of IoT to Improve Parcel Delivery System</b>	<b>497</b>
<i>Ha Yoon Song, Hyo Chang Han</i>	
<b>On Coverage of 3D Terrains by Wireless Sensor Networks</b>	<b>501</b>
<i>Mostefa Zafer, Mustapha Reda Senouci, Mohamed Aissani</i>	
<b>Smart Urban Objects to Enhance Safe Participation in Major Events for the Elderly</b>	<b>505</b>
<i>Tobias Zimpel, Marvin Hubl</i>	

---

## INFORMATION SYSTEMS AND TECHNOLOGY

---

<b>Call For Papers</b>	<b>515</b>
------------------------	------------

---

### 17<sup>TH</sup> CONFERENCE ON ADVANCED INFORMATION TECHNOLOGIES FOR MANAGEMENT

---

<b>Call For Papers</b>	<b>517</b>
<b>Factors Influencing the Intended Adoption of Digital Transformation: A South African Case Study</b>	<b>519</b>
<i>Jean-Paul Van Belle, Rion van Dyk</i>	
<b>Aspects of Mobility of e-Marketing from Customer Perspective</b>	<b>529</b>
<i>Witold Chmielarz, Marek Zborowski, Üyesi Mesut Atasever</i>	

<b>Network Effects in Online Marketplaces: The Case of Kiva</b>	<b>535</b>
<i>Haim Mendelson, Yuanyuan Shen</i>	
<b>The Approach to Applications Integration for World Data Center Interdisciplinary Scientific Investigations</b>	<b>539</b>
<i>Grzegorz Nowakowski, Sergii Telenyk, Kostiantyn Yefremov, Volodymyr Khmeliuk</i>	
<b>Information Systems Development and Usage with Consideration of Privacy and Cyber Security Aspects</b>	<b>547</b>
<i>Janusz Jabłoński, Silva Robak</i>	
<b>Deriving Workflow Privacy Patterns from Legal Documents</b>	<b>555</b>
<i>Marcin Robak, Erik Buchmann</i>	
<b>Using Blockchain to Access Cloud Services: A Case of Financial Service Application</b>	<b>565</b>
<i>Min-Han Tseng, Shuchih Ernest Chang, Tzu-Yin Kuo</i>	
<b>Predicting Automotive Sales using Pre-Purchase Online Search Data</b>	<b>569</b>
<i>Philipp Wachter, Tobias Widmer, Achim Klein</i>	
<b>Exploring Levels of ICT Adoption and Sustainable Development – The Case of Polish Enterprises</b>	<b>579</b>
<i>Ewa Ziemia</i>	
<hr/>	
<b>1<sup>ST</sup> SPECIAL SESSION ON DATA SCIENCE IN HEALTH</b>	
<hr/>	
<b>Call For Papers</b>	<b>589</b>
<b>Medical data exploration based on the heterogeneous data sources aggregation system</b>	<b>591</b>
<i>Andrzej Opaliński, Krzysztof Regulski, Barbara Mrzygłód, Mirosław Głowacki, Aleksander Kania, Paweł Nastalek, Natalia Celejewska-Wójcik, Grażyna Bochenek, Krzysztof Stadek</i>	
<b>Mapping of Dental Care in the Czech Republic: Case Study of Graduates Distribution in Practice</b>	<b>599</b>
<i>Matěj Karolyi, Jakub Šcavnický, Jan Bud'a, Tereza Jurková, Monika Mazalová, Martin Komenda</i>	
<b>Medical prescription classification: a NLP-based approach</b>	<b>605</b>
<i>Vincenza Carchiolo, Alessandro Longheu, Giuseppa Reitano, Luca Zagarella</i>	
<hr/>	
<b>1<sup>ST</sup> WORKSHOP ON DATA ANALYSIS AND COMPUTATION FOR DIGITAL ECOSYSTEMS</b>	
<hr/>	
<b>Call For Papers</b>	<b>611</b>
<b>Location Intelligence in Cogenerated Heating Potential Data</b>	<b>613</b>
<i>Almir Karabegovic, Mirza Ponjavic, Neven Duic, Tomislav Novosel</i>	
<b>MOPS – A feasibility Study for working with GPS and sensor data in a medical context</b>	<b>621</b>
<i>Christof Meigen, Mandy Vogel, Jan Bumberger</i>	
<hr/>	
<b>14<sup>TH</sup> CONFERENCE ON INFORMATION SYSTEMS MANAGEMENT</b>	
<hr/>	
<b>Call For Papers</b>	<b>625</b>
<b>Exploring Determinants of M-Government Services: A Study from the Citizens' Perspective in Saudi Arabia</b>	<b>627</b>
<i>Mohammed Alonazi, Natalia Beloff, Martin White</i>	
<b>Developing a Model and Validating an Instrument for Measuring the Adoption and Utilisation of Mobile Government Services Adoption in Saudi Arabia</b>	<b>633</b>
<i>Mohammed Alonazi, Natalia Beloff, Martin White</i>	
<b>Towards Data Quality Runtime Verification</b>	<b>639</b>
<i>Janis Bicevskis, Zane Bicevska, Anastasija Nikiforova, Ivo Oditis</i>	
<b>BPM Tools for Asset Management in Renewable Energy Power Plants</b>	<b>645</b>
<i>Vincenza Carchiolo, Giovanni Catalano, Michele Malgeri, Carlo Pellegrino, Giulio Platania, Natalia Trapani</i>	

<b>Identification of Heuristics for Assessing the Usability of Websites of Public Administration Units</b>	<b>651</b>
<i>Helena Dudycz, Łukasz Krawiec</i>	
<b>Motivations for BPM Adoption: Initial Taxonomy based on Online Success Stories</b>	<b>659</b>
<i>Renata Gabryelczyk, Aneta Biernikowicz</i>	
<b>Multi-criteria approach to viral marketing campaign planning in social networks, based on real networks, network samples and synthetic networks</b>	<b>663</b>
<i>Artur Karczmarczyk, Jarosław Jankowski, Jarosław Wątróbski</i>	
<b>An Approach to Customer Community Discovery</b>	<b>675</b>
<i>Jerzy Korczak, Maciej Pondel, Wiktor Sroka</i>	
<b>ICT Usage in Industrial Symbiosis: Problem Identification and Study Design</b>	<b>685</b>
<i>Linda Kosmol, Christian Leyh</i>	
<b>On the Use of Predictive Models for Improving the Quality of Industrial Maintenance: An Analytical Literature Review of Maintenance Strategies</b>	<b>693</b>
<i>Oana Merkt</i>	
<b>Visual Rule Editor for E-Guide Gamification Web Platform</b>	<b>705</b>
<i>Jakub Swacha, Artur Kulpa, Karolina Muszyńska</i>	
<b>A Design and Experiment of Automation Management System for Platform as a Service</b>	<b>711</b>
<i>Alalaa Tashkandi</i>	
<b>Project Management Tasks in Agile Projects: A Quantitative Study</b>	<b>717</b>
<i>Gloria Miller</i>	

---

## **25<sup>TH</sup> CONFERENCE ON KNOWLEDGE ACQUISITION AND MANAGEMENT**

---

<b>Call For Papers</b>	<b>723</b>
<b>Analysis of Relationship between Personal Factors and Visiting Places using Random Forest Technique</b>	<b>725</b>
<i>Young Myung Kim, Ha Yoon Song</i>	
<b>Automated Generation of Business Process Models using Constraint Logic Programming in Python</b>	<b>733</b>
<i>Tymoteusz Paszun, Piotr Wiśniewski, Krzysztof Kluza, Antoni Ligeza</i>	
<b>Analysis of the Correlation Between Personal Factors and Visiting Locations With Boosting Technique</b>	<b>743</b>
<i>Ha Yoon Song, Yun JiSeon</i>	
<b>Do online reviews reveal mobile application usability and user experience? The case of WhatsApp</b>	<b>747</b>
<i>Paweł Weichbroth, Anna Baj-Rogowska</i>	
<b>Parameter Setting Problem in the Case of Practical Vehicle Routing Problems with Realistic Constraints</b>	<b>755</b>
<i>Emir Žunić, Dženana Đonko</i>	

---

## **SOFTWARE AND SYSTEM ENGINEERING**

---



---

### **JOINT 39<sup>TH</sup> IEEE SOFTWARE ENGINEERING WORKSHOP (SEW-39) AND 6<sup>TH</sup> INTERNATIONAL WORKSHOP ON CYBER-PHYSICAL SYSTEMS (IWCPS-6)**

---

<b>Call For Papers</b>	<b>761</b>
<b>Handling of Categorical Data in Software Development Effort Estimation: A Systematic Mapping Study</b>	<b>763</b>
<i>Fatima Azzahra Amzal, Ali Idri</i>	
<b>Big Data Platform for Smart Grids Power Consumption Anomaly Detection</b>	<b>771</b>
<i>Peter Lipcák, Martin Macak, Bruno Rossi</i>	

<b>Search for the Memory Duplicities in the Java Applications Using Shallow and Deep Object Comparison</b>	<b>781</b>
<i>Richard Lipka, Tomas Potuzak</i>	
<b>Redesigning Method Engineering Education Through a Trinity of Blended Learning Measures</b>	<b>791</b>
<i>Sietse Overbeek, Sjaak Brinkkemper</i>	

---

### **3<sup>RD</sup> INTERNATIONAL CONFERENCE ON LEAN AND AGILE SOFTWARE DEVELOPMENT**

---

<b>Call For Papers</b>	<b>799</b>
<b>Preliminary Citation and Topic Analysis of International Conference on Agile Software Development Papers (2002-2018)</b>	<b>803</b>
<i>Muhammad Ovais Ahmad, Paivi Raulamo-Jurvanen</i>	
<b>Factors that contribute significantly to Scrum adoption</b>	<b>813</b>
<i>Ridewaan Hanslo, Ernest Mnkandla, Anwar Vaheed</i>	
<b>Create your own agile methodology for your research and development team</b>	<b>823</b>
<i>Enikő Ilyés</i>	
<b>Real-Life Challenges in Automotive Release Planning</b>	<b>831</b>
<i>Kristina Marner, Sven Theobald, Stefan Wagner</i>	
<b>On the Agile Mindset of an Effective Team – An Industrial Opinion Survey</b>	<b>841</b>
<i>Jakub Miler, Paulina Gaida</i>	
<b>Scaling agile on large enterprise level – systematic bundling and application of state of the art approaches for lasting agile transitions</b>	<b>851</b>
<i>Alexander Poth, Mario Kottke, Andreas Riel</i>	
<b>Participating in an Industry Based Social Service Program: a Report of Student Perception of What They Learn and What They Need</b>	<b>861</b>
<i>Miguel Hécatl Morales Trujillo, Gabriel Alberto García Mireles</i>	
<b>Playing the Sprint Retrospective</b>	<b>871</b>
<i>Maciej Wawryk, Yen Ying Ng</i>	
<b>Security-oriented agile approach with AgileSafe and OWASP ASVS</b>	<b>875</b>
<i>Katarzyna Łukasiewicz, Sara Cygańska</i>	

---

### **7<sup>TH</sup> CONFERENCE ON MULTIMEDIA, INTERACTION, DESIGN AND INNOVATION**

---

<b>Call For Papers</b>	<b>879</b>
<b>Exploration of older drivers interaction with conversation assistant</b>	<b>881</b>
<i>Jakub Berka, Lukas Chvatal, Zdenek Mikovec</i>	
<b>Supporting personalized care of older adults with vision and cognitive impairments by user modeling</b>	<b>891</b>
<i>Miroslav Macik, Petr Bilek, Zdenek Mikovec</i>	
<b>Gamified Augmented Reality Training for An Assembly Task: A Study About User Engagement</b>	<b>901</b>
<i>Diep Nguyen, Gerrit Meixner</i>	

---

### **6<sup>TH</sup> DOCTORAL SYMPOSIUM ON RECENT ADVANCES IN INFORMATION TECHNOLOGY**

---

<b>Call For Papers</b>	<b>905</b>
<b>Sustainable Management of Marine Fish Stocks by Means of Sliding Mode Control</b>	<b>907</b>
<i>Katharina Benz, Claus Rech, Paolo Mercorelli</i>	
<b>An effective industrial control approach</b>	<b>911</b>
<i>Michal Kostolani, Justín Murín, Štefan Kozák</i>	

<b>PN2ARDUINO - A New Petri Net Software Tool For Control Of Discrete-event And Hybrid Systems Using Arduino Microcontrollers</b>	<b>915</b>
<i>Erik Kučera, Oto Haffner, Roman Leskovský</i>	
<b>Use of Holographic Technology in Online Experimentation</b>	<b>921</b>
<i>Jakub Matišák, Matej Rábek, Katarína Žáková</i>	
<b>Proposal of Mechatronic Devices Control using Mixed Reality</b>	<b>925</b>
<i>Erich Stark, Erik Kučera, Peter Drahoš, Oto Haffner</i>	
<b>The Effects of Augmented Training Dataset on Performance of Convolutional Neural Networks in Face Recognition System</b>	<b>929</b>
<i>Mehmet Ali Kutlugün, Yahya Şirin, Mehmet Ali Karakaya</i>	
<b>Author Index</b>	<b>933</b>



# 14<sup>th</sup> International Symposium Advances in Artificial Intelligence and Applications

**A** AIA'19 brings together scientists and practitioners to discuss their latest results and ideas in all areas of Artificial Intelligence. We hope that successful applications presented at AAIA'19 will be of interest to researchers who want to know about both theoretical advances and latest applied developments in AI.

## TOPICS

Papers related to theories, methodologies, and applications in science and technology in the field of AI are especially solicited. Topics covering industrial applications and academic research are included, but not limited to:

- Decision Support
- Machine Learning
- Fuzzy Sets and Soft Computing
- Rough Sets and Approximate Reasoning
- Data Mining and Knowledge Discovery
- Data Modeling and Feature Engineering
- Data Integration and Information Fusion
- Hybrid and Hierarchical Intelligent Systems
- Neural Networks and Deep Learning
- Bayesian Networks and Bayesian Reasoning
- Case-based Reasoning and Similarity
- Web Mining and Social Networks
- Business Intelligence and Online Analytics
- Robotics and Cyber-Physical Systems
- AI-centered Systems and Large-Scale Applications

## PROFESSOR ZDZISŁAW PAWLAK BEST PAPER AWARDS

We are proud to continue the tradition started at the AAIA'06 and grant two "Professor Zdzisław Pawlak Best Paper Awards" for contributions which are outstanding in their scientific quality. The two award categories are:

- Best Student Paper. Papers qualifying for this award must be marked as "Student full paper" to be eligible.
- Best Paper Award.

Each award carries a prize of 300 EUR funded by the Mazowsze Chapter of the Polish Information Processing Society.

## EVENT CHAIRS

- **Kwaśnicka, Halina**, Wrocław University of Science and Technology, Poland
- **Markowska-Kaczmar, Urszula**, Wrocław University of Science and Technology, Poland

## ADVISORY BOARD

- **Kacprzyk, Janusz**, Polish Academy of Sciences, Poland
- **Marek, Victor**, University of Kentucky, United States
- **Matwin, Stan**, Dalhousie University, Canada
- **Michalewicz, Zbigniew**, University of Adelaide, Australia
- **Skowron, Andrzej**, University of Warsaw, Poland
- **Ślęzak, Dominik**, University of Warsaw, Poland

## TRACK PROGRAM COMMITTEE

- **Derksen, Christian**, SEN-MAS'19
- **Lasek, Piotr**, AIMA'19
- **Loukanova, Roussanka**, AIRLangComp'19
- **Markowska-Kaczmar, Urszula**, AAIA'19
- **Mozgovoy, Maxim**, ASIR'19
- **Zaharie, Daniela**, WCO'19

## PROGRAM COMMITTEE

- **AbdelRaouf, Ashraf**, Misr International University, Egypt
- **Abiyev, Rahib**, Near East University, Turkey
- **Al-Mardini, Mamoun**, University of Florida, United States
- **Alirezaie, Javad**
- **Antonelli, Michela**
- **Baron, Grzegorz**
- **Bartkowiak, Anna**, Wrocław University, Poland
- **Bembenik, Robert**
- **Betliński, Paweł**, Security On Demand, Poland
- **Boryczka, Mariusz**
- **Bouguelia, Mohamed-Rafik**
- **Bryniarska, Anna**, Opole University of Technology, Poland
- **Błaszczyszki, Jerzy**, Poznań University of Technology, Poland
- **Calpe, Javier**
- **Camasta, Francesco**
- **Chakraverty, Shampa**, Netaji Subhas Institute of Technology, India
- **Chu, Henry**, University of Louisiana at Lafayette, United States
- **do Carmo Nicoletti, Maria**, UFSCar & FACCAMP, Brazil
- **Franova, Marta**, CNRS, LRI & INRIA, France
- **Froelich, Wojciech**, University of Silesia, Poland
- **Gawrysiak, Piotr**
- **Girardi, Rosario**, UNIRIO, Brazil
- **Goyal, Anuj**
- **Jaromczyk, Jerzy**, University of Kentucky, United States

- **Jatowt, Adam**, Kyoto University, Japan
- **Jin, Xiaolong**, Institute of Computing Technology, Chinese Academy of Sciences, China
- **Kasprzak, Włodzimierz**, Warsaw University of Technology, Poland
- **Korbicz, Józef**, University of Zielona Góra, Poland
- **Krol, Dariusz**
- **Kryszkiewicz, Marzena**, Warsaw University of Technology, Poland
- **Kulikowski, Juliusz**, Institute of Biocybernetics and Biomedical Engineering, Poland
- **Kwiatkowski, Jan**
- **Labati, Ruggero Donida**
- **Lewis, Rory**, University of Colorado Colorado Springs, United States
- **Lim, Chee Peng**
- **Matson, Eric T.**, Purdue University, United States
- **Menasalvas, Ernestina**, Universidad Politécnica de Madrid, Spain
- **Moshkov, Mikhail**, King Abdullah University of Science and Technology, Saudi Arabia
- **Myszkowski, Paweł B.**, Wrocław University of Technology, Poland
- **Nowostawski, Mariusz**, Norwegian University of Technology and Science (NTNU), Norway
- **Ohsawa, Yukio**, University of Tokyo, Japan
- **Olech, Łukasz**, Wrocław University of Science and Technology, Poland
- **Peters, Georg**, Munich University of Applied Sciences, Germany
- **Po, Laura**, Università di Modena e Reggio Emilia, Italy
- **Porta, Marco**, University of Pavia, Italy
- **Przewoźniczek, Michał**, Wrocław University of Technology / MP2 company, Poland
- **Przybyła-Kasperek, Małgorzata**, University of Silesia, Poland
- **Raś, Zbigniew**, University of North Carolina at Charlotte, United States
- **Rauch, Jan**, University of Economics, Prague, Czech Republic
- **Reformat, Marek**, University of Alberta, Canada
- **Sas, Jerzy**, Wrocław University of Technology, Poland
- **Schaefer, Gerald**, Loughborough University, United Kingdom
- **Sikora, Marek**, Silesian University of Technology, Poland
- **Sikos, Leslie F.**, University of South Australia, Australia
- **Skonieczny, Lukasz**
- **Śluzek, Andrzej**, Khalifa University, United Arab Emirates
- **Stańczyk, Urszula**, Silesian University of Technology, Poland
- **Subbotin, Sergey**, Zaporizhzhya National Technical University, Ukraine
- **Sydow, Marcin**, Polish Academy of Sciences & Polish-Japanese Academy of Information Technology, Poland
- **Szczęch, Izabela**, Poznań University of Technology, Poland
- **Szczuka, Marcin**, University of Warsaw, Poland
- **Szpakowicz, Stan**, University of Ottawa, Canada
- **Szwed, Piotr**, AGH University of Science and Technology, Poland
- **Tarnowska, Katarzyna**, San Jose State University, United States
- **Tomczyk, Arkadiusz**, Łódź University of Technology, Poland
- **Unland, Rainer**, Universität Duisburg-Essen, Germany
- **Unold, Olgierd**, Wrocław University of Technology, Poland
- **Wróblewska, Anna**, Warsaw University of Technology, Poland
- **Zakrzewska, Danuta**, Łódź University of Technology, Poland
- **Zatwarnicka, Anna**
- **Zatwarnicki, Krzysztof**
- **Zielosko, Beata**, University of Silesia, Poland

# Clash Royale Challenge: How to Select Training Decks for Win-rate Prediction

Andrzej Janusz\*<sup>†</sup>, Łukasz Grad\*<sup>†</sup>, Marek Grzegorowski\*<sup>†</sup>

\*Institute of Informatics, University of Warsaw, Poland

<sup>†</sup>Esensei Sp. z o.o., Poland

Contact Email: janusza@mimuw.edu.pl

**Abstract**—We summarize the sixth data mining competition organized at the Knowledge Pit platform in association with the Federated Conference on Computer Science and Information Systems series, titled Clash Royale Challenge: How to Select Training Decks for Win-rate Prediction. We outline the scope of this challenge and briefly present its results. We also discuss the problem of acquiring knowledge about new notions from video games through an active learning cycle. We explain how this task is related to the problem considered in the challenge and share results of experiments that we conducted to demonstrate usefulness of the active learning approach in practice.

**Keywords**—Data Mining Contest; Training Subset Selection; Win-rates Prediction; Active Learning; Clash Royale

## I. INTRODUCTION

Video games, and especially mobile games, are considered as one of the domains in which a huge amount of data is generated by players on a daily basis. Utilization of such data in practical applications requires complex analysis towards a proper understanding of hidden concepts and time-consuming data preparation process. In particular, it is often necessary to provide labels of data records that we want to use for model training. Even though it is very laborious, this process is necessary to train intelligent models that could provide value to end-users. Due to limited time and budget, it is usually possible to label only a small amount of data. The “as-is” market standard is to manually label data records. To handle this, a number of corporations utilize crowd-computing services to outsource data-labeling capability. However, it seems that the labeling process could be optimized using approaches related to active learning (AL) [1]. An alternative way could be, so-called weak supervision, where less reliable labels are generated using simple heuristics using domain knowledge [2].

In this research, we use as an example a popular mobile collectible card video game – Clash Royale – which combines elements of collectible card game and tower defense genres (<https://clashroyale.com/>). In this game, players build decks consisting of 8 cards that represent playable troops, buildings, and spells, which they use to attack opponent’s towers and defend against their cards. Using good decks is one of the critical abilities of successful Clash Royale players. We describe a challenge in which we take on a problem of measuring and predicting the deck effectiveness in 1v1 ladder games. In particular, we would like to find out whether it is possible to train an efficient win-rate prediction model on a relatively

small subset of decks, whose win-rates were estimated in the past. Such a task can also be considered in the context of active learning, as a selection of a data batch that should be labeled and used for training a win-rate prediction model.

The remaining of the paper is organized as follows: In Section II, we discuss a context for the competition, i.e. the problem of active learning from video game data. In Section III, we briefly describe the competition and summarize its results. In Section IV, we present a framework for predicting win-rates of Clash Royale decks. In Section V, we conclude the paper and draw some directions for future research.

## II. ACTIVE LEARNING FROM VIDEO GAME DATA

Active learning is a domain within the field of machine learning, in which the learning algorithm can interactively query an oracle about labels (or more generally, target attribute values) of some limited number of training records [3]. Its applications are particularly suitable when the availability of labeled data is limited. In such cases, to train reliable prediction models, it is often necessary to perform a laborious and costly process of manual data labeling. Through the use of AL, it is possible to facilitate this process by allowing the algorithm to choose records which seem the most beneficial for learning [4]. Such a selection of training examples is performed based on results of a model constructed in a previous iteration of the AL cycle (Figure 1). The importance of unlabeled examples is determined by the confidence of their classification or by the expected model change after including the instances to the training data [5]. To deal with the cold-start problem, the first training batch is typically selected at random or by using some clustering technique to find a diverse yet representative set of initial examples for labeling [6]. Then, in subsequent iterations of the AL cycle, additional examples are selected and the prediction model is continuously improved [7].

In practice, the data is usually labeled by a committee of experts and the oracle is implemented as a voting system. Since humans are prone to errors, several experts assign labels to each data record, and the final labeling is determined by voting [1]. Research in the AL field focus mainly on algorithms for selecting a single data record for labeling in each iteration of the AL cycle. However, when there are many available experts, it is more efficient to choose larger batches. In this way, experts who label faster do not have to wait until others finish their tasks and the prediction model is updated.

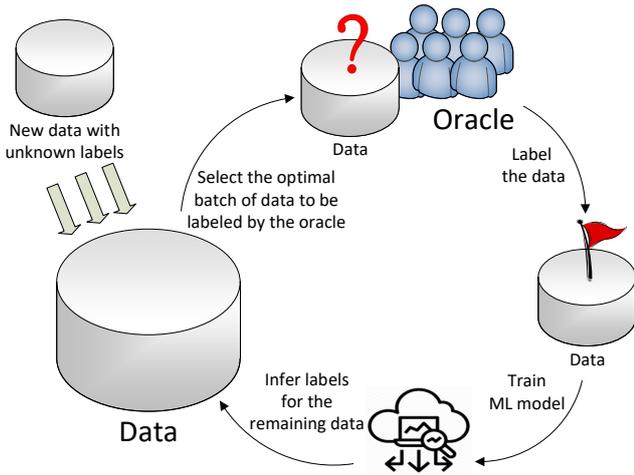


Fig. 1. An active learning cycle. The oracle is interactively queried about labels of records which are selected as the most beneficial for learning by the algorithm.

In a context of video game data, the role of experts can be assumed by the community of players. As a consequence, the committee which assigns labels can be quite large and diverse. This fact impacts the active learning setup in two main aspects:

- 1) In order to optimize the efficiency of the AL cycle and avoid lags in the labeling process, algorithms need to select many data records for labeling at a time.
- 2) Each of selected records should be shown to a subset of available labelers. The voting algorithm should take into account the diversity of labelers and remain robust, even in a presence of a large number of noisy labels.

The system governing the AL cycle should be able to guarantee that whenever there is an available labeler, it can provide a new example for labeling. Moreover, the oracle should be able to find a consensus among contradicting assignments of labels and be able to discard those whose quality is likely to be low. This can be done through the estimation of labelers' expertise, combined with a weighted voting schema [8].

### III. CLASH ROYALE CHALLENGE

The task in Clash Royale Challenge was related to the first of the two problems mentioned in Section II, namely, the selection of a data subset that allows to construct an efficient model for predicting win-rates of Clash Royale decks. The competition took place between April 24, 2019 and June 12, 2019, under the auspices of 14<sup>th</sup> Federated Conference on Computer Science and Information Systems. It was organized on the KnowledgePit platform which underwent a significant lift-up shortly before the start of the challenge (<https://knowledgepit.ml/clash-royale-challenge/>).

The competition's task could also be viewed as a continuation of the topic started in the previous year, i.e. the prediction of win-rates of decks from collectible card video games [9]. The ability to assess quality of decks in a continuously

evolving game is one of core features of an advisory system for players, called SENSEI, which is being developed by one of the competition's sponsors [10].

Data in this challenge consisted of 100.000 Clash Royale decks that were most commonly used by players during three consecutive league seasons in 1v1 ladder games. They were provided in a tabular format. Each row the training set corresponded to a Clash Royale deck and was described by four columns. The first one listed eight cards that constitute the deck. The second and third column showed the number of games played with the deck, and the number of players that were using it, respectively. These values were computed based on over 160.000.000 game results obtained using the RoyaleAPI service (<https://royaleapi.com/>) and SENSEI's data acquisition module. The last column indicated estimations of win-rates of the decks, that were calculated based on games played in the given time window. Participants were asked to indicate ten subsets of those decks, with sizes fixed to 600, 700, . . . , 1500. These subsets were ought to allow training efficient support vector regression models (SVR) with radial kernels [11] for a purpose of win-rate prediction (one model for each training data subset). Competitors could also tune hyper-parameters of the models.

#### A. Evaluation of results and participation in the challenge

The quality of solutions was assessed by measuring the prediction performance of the models trained on data subsets indicated by the participants. This evaluation step was conducted on a separate set of decks. This test data consisted of decks that were popular during the three game seasons after the training data period. This set was not revealed to participants before the end of the challenge. However, a small subset of decks from the test period (a validation data set) was given to participants. It is also worth noticing that the same decks could appear in both the training and evaluation data, but they were likely to have different win-rates. The cause of those differences is the fact that the game evolves in time, players adapt to new strategies, and the balance of individual cards (and their popularity) changes from one season to another.

During the competition, submitted solutions were evaluated online, and the preliminary results were published on Leaderboard. The preliminary score was computed on a randomly selected set of 2000 test records, fixed for all participants. The final evaluation was performed after completion of the competition using the remaining part of the test data. Each teams was oblige to submit a report describing their approach before the end of the challenge.

The measure chosen for the assessment of solutions was the R-squared. If we denote a prediction for a test instance  $i$  as  $f_i$ , and its reference win-rate as  $y_i$ , the R-squared metric is:

$$R^2 = 1 - \frac{RSS}{TSS} \quad (1)$$

where  $RSS$  and  $TSS$  are the residual and total sum of squares, respectively:

$$RSS = \sum_i (y_i - f_i)^2, \quad TSS = \sum_i (y_i - \bar{y})^2$$

TABLE I. FINAL R-SQUARED VALUES AND NUMBER OF SUBMISSIONS FROM TOP-RANKED TEAMS. THE LAST ROW SHOWS THE RESULT OBTAINED BY THE BASELINE SOLUTION.

team name	rank	number of submissions	final result
Dymitr	1	144	0.2552
amy	2	123	0.2530
ru	3	25	0.2257
ms	4	51	0.2241
--	5	30	0.2215
...	...	...	...
baseline	14	1	0.1564

and  $\bar{y} = \frac{1}{N} \sum_i y_i$ .

A value of this metric was computed independently for predictions made by SVR models trained on each of the subsets indicated in the submitted solutions. The final score was an average of the obtained results.

**B. Summary of the competition results**

The scores obtained by top-ranked teams are presented in Table I. The baseline in this challenge was obtained using a simple algorithm that utilizes basic properties of the SVR model, i.e., only records which correspond to the support vectors have any impact on the model. A  $\nu$ -regression SVR was trained on a subset of the most popular training decks with the parameter values set such that the number of selected support vectors corresponded the the desired sizes of target sets. These vectors were taken as the baseline solution.

Participants of the challenge were able to significantly improve over the baseline score. Unfortunately, no team from the top 10 was using an approach that could be applied to the considered problem in practice. The winners were using a greedy search heuristic to limit the candidate decks. Then, they fine-tuned the final sets using exhaustive search. In both cases, the quality of fit was computed as the R-squared value obtained on the validation data. In practice, such data would not be available. Thus any supervised search heuristic would not be feasible. More detailed description of the winning approach can be found in [12]. In Section IV, we propose an alternative method which solves the competition problem without a need for a validation sample. It uses an approach inspired by active learning and can be utilized in a way similar to the AL cycle to continuously adapt to a changing game.

**IV. ESTIMATION OF WIN-RATES USING LIMITED DATA**

We approach the problem of win-rate estimation in Clash Royale using limited training data from a pool-based active learning perspective. Specifically, we propose a solution based on density weighted batch uncertainty sampling. For uncertainty sampling, we provide an informativeness function tailored to the case of known, but noisy labels. Such an approach is viable in the context of win-rate prediction because they change in time due to balance changes in the game. Even though we can always estimate win-rates using historical data (e.g. data from a previous game season), such estimates are likely to be invalid for new game seasons.

**A. An informativeness measure**

Formally, given a training data set  $T$  consisting of records  $(x_i, y_i)_{i=1}^N$  with known label noise  $Var[y_i] = \sigma_i^2$  and a model  $M$ , we search for a training subset of given size  $K$ , such that the model trained on this subset achieves the lowest generalization error, i.e.:

$$A^* = \operatorname{argmin}_{A:|A|=K} \mathbb{E}_{(X,Y)} [l(Y, f_A^M(X))] \tag{2}$$

where  $f_A^M$  is the mapping induced by the model  $M$  trained on subset  $A$  and  $l$  is the mean squared error loss function.

In our method, we begin by choosing an initial training subset  $A_0$  of size  $m$  at random. Then, at each step, we greedily select a sample that maximizes the importance:

$$x^* = \operatorname{argmax}_{x:T} [\phi(x)^\alpha \times Sim(x)^\beta \times Dis(x)^\gamma] \tag{3}$$

where  $\phi$  measures the informativeness of samples,  $Sim(x) = (\frac{1}{u} \sum_{i=1}^u sim(x, x_i))$  is a measure of a representativeness, and  $Dis(x) = (\frac{1}{b} \sum_{i=1}^b dis(x, x_i^B))$  measures the dissimilarity in the current batch, assuming that we have already chosen samples  $(x_0^B, \dots, x_b^B)$ . Parameters  $\alpha, \beta, \gamma$  control the relative importance of each factor. In this work, we set each of the parameters to 1. Similarity measure used in all our experiments was the Jaccard index:  $sim(x_1, x_2) = \frac{|x_1 \cap x_2|}{|x_1 \cup x_2|}$  and  $dis(x_1, x_2) = 1 - sim(x_1, x_2)$ .

To derive the measure of informativeness  $\phi$ , we assume normality of the response  $Y$ . Given a trained model  $M$  and the posterior distribution of the response  $g_M(X_i) \sim \mathcal{N}(\mu_i, \tau_i^2)$ , we obtain the posterior predictive distribution  $\hat{Y}_i \sim \mathcal{N}(\mu_i, \tau_i^2 + \sigma_i^2)$ . This is valid, since noise process and the posterior distribution are independent Gaussians. Now, given a sample  $(x_i, y_i)$  we define the informativeness as

$$\begin{aligned} \phi(x_i) &= 1 - \mathbb{P}(|\hat{Y}_i - \mu_i| > |d_i|) \\ &= 1 - \mathbb{P}((\hat{Y}_i - \mu_i) > |d_i|) - \mathbb{P}((\hat{Y}_i - \mu_i) < -|d_i|) \\ &= 1 - \mathbb{P}\left(Z_i > \frac{|d_i|}{\sqrt{\sigma_i^2 + \tau_i^2}}\right) - \mathbb{P}\left(Z_i < \frac{-|d_i|}{\sqrt{\sigma_i^2 + \tau_i^2}}\right) \\ &= 1 - (1 - \Phi(|d_i|)) - \Phi(-|d_i|) \\ &= 1 - 2\Phi(-|d_i|) \end{aligned}$$

where  $d_i = y_i - \mu_i$ ,  $Z_i$  denotes a standard normal variable and  $\Phi$  is a standard normal CDF.

In our experiments, we used Gaussian Process Regression model [13], along with absolute exponential covariance kernel:

$$K_{\kappa^2, \lambda}(x, x^*) = \kappa^2 \exp\left(-\frac{|x - x^*|}{\lambda}\right) \tag{4}$$

where  $\kappa^2$  and  $\lambda$  are kernel hyper-parameters optimized during model fitting.

**B. Experimental results**

We compared the results obtained using our approach to the best solution from the winners of the challenge [12]. Instead of computing the R-squared metric, we simply measured the root mean squared error (RMSE) of SVR models trained on

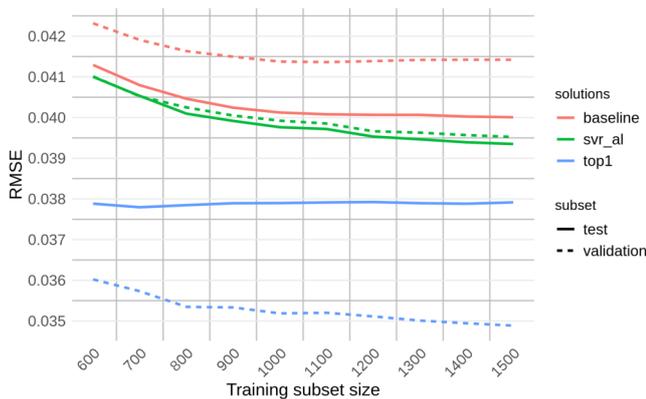


Fig. 2. Results of the compared training subset selection methods. The dotted lines indicate RMSE values obtained on the validation data, whereas the solid lines correspond to the final test set.

each data subset from the winner’s solution, and on subsets of corresponding sizes found using our method. We checked errors of the trained models on the final test set from the challenge, as well as on the validation set which was fully available to competing teams during the competition. Figure 2 shows those results. To provide a better reference, we also computed RMSE values achieved by the baseline method.

Even though our method achieved lower scores than the winner’s for all subset sizes, it is important to notice that in practice, when the validation set is not available, it would be much more useful. For the largest subset size the difference between is lower than 0.002, which seems negligible considering the variance of predicted win-rates. Furthermore, the AL-inspired method is always better than the baseline. The plot also shows that our method is not over-fitted to any particular data subset, whereas the winners achieved much better results on the validation set than on the final test set. Interestingly, RMSE of the winning solution on the final set does not decrease with the growing size of training data subset (it is even slightly higher). This could be regarded as another argument in favour of our method.

## V. SUMMARY

In this paper, we described Clash Royale Challenge organized at the KnowledgePit platform, whose scope was on finding an optimal data subset for training win-rate prediction models. Our competition attracted 115 teams from 18 countries. Among the participating teams, 68 submitted at least one solution file which was ranked on the public Leaderboard. More than 40 of those teams decided to disclose their approach by uploading short reports.

A dominating approach used by competitors was based on a greedy search heuristic with a fit function that used an additional validation set. As an alternative, we proposed a method inspired by active learning, which is more suitable to solve the considered problem in practical applications. In the presented experiments, we showed that it can be effective. In the future, we will focus on extending our approach in the

context of large batch sampling, e.g. by effectively utilizing a predictive covariance matrix for computing the informativeness function utilized by our method.

## ACKNOWLEDGMENTS

This research was co-funded by Smart Growth Operational Programme 2014-2020, financed by European Regional Development Fund under GameINN project POIR.01.02.00-00-0184/17, operated by National Centre for Research and Development in Poland.

## REFERENCES

- [1] S. Yan, K. Chaudhuri, and T. Javidi, “Active learning from imperfect labelers,” in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, ser. NIPS’16. USA: Curran Associates Inc., 2016, pp. 2136–2144. [Online]. Available: <http://dl.acm.org/citation.cfm?id=3157096.3157335>
- [2] S. H. Bach, D. Rodriguez, Y. Liu, C. Luo, H. Shao, C. Xia, S. Sen, A. Ratner, B. Hancock, H. Alborzi, R. Kuchhal, C. Ré, and R. Malkin, “Snorkel drybell: A case study in deploying weak supervision at industrial scale,” in *SIGMOD Conference*. ACM, 2019, pp. 362–375.
- [3] B. Settles, *Active Learning*. Morgan & Claypool Publishers, 2012.
- [4] E. Lughofer, “Hybrid active learning for reducing the annotation effort of operators in classification systems,” *Pattern Recogn.*, vol. 45, no. 2, pp. 884–896, Feb. 2012. [Online]. Available: <http://dx.doi.org/10.1016/j.patcog.2011.08.009>
- [5] W. Cai, Y. Zhang, Y. Zhang, S. Zhou, W. Wang, Z. Chen, and C. Ding, “Active learning for classification with maximum model change,” *ACM Trans. Inf. Syst.*, vol. 36, no. 2, pp. 15:1–15:28, Aug. 2017. [Online]. Available: <http://doi.acm.org/10.1145/3086820>
- [6] H. T. Nguyen and A. Smeulders, “Active learning using pre-clustering,” in *Proceedings of the Twenty-first International Conference on Machine Learning*, ser. ICML ’04. New York, NY, USA: ACM, 2004, pp. 79–. [Online]. Available: <http://doi.acm.org/10.1145/1015330.1015349>
- [7] K. Konyushkova, R. Sznitman, and P. Fua, “Learning active learning from real and synthetic data,” *CoRR*, vol. abs/1703.03365, 2017. [Online]. Available: <http://arxiv.org/abs/1703.03365>
- [8] C. Zhang and K. Chaudhuri, “Active learning from weak and strong labelers,” in *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1*, ser. NIPS’15. Cambridge, MA, USA: MIT Press, 2015, pp. 703–711. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2969239.2969318>
- [9] A. Janusz, T. Tajmajer, M. Świechowski, Ł. Grad, J. Puczniewski, and D. Ślęzak, “Toward an intelligent HS deck advisor: Lessons learned from aiaa’18 data mining competition,” in *Proceedings of the 2018 Federated Conference on Computer Science and Information Systems, FedCSIS 2018, Poznań, Poland, September 9-12, 2018.*, M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds., 2018, pp. 189–192. [Online]. Available: <https://doi.org/10.15439/2018F386>
- [10] A. Janusz, D. Ślęzak, S. Stawicki, and K. Stencel, “SENSEI: an intelligent advisory system for the esports community and casual players,” in *2018 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2018, Santiago, Chile, December 3-6, 2018*. IEEE Computer Society, 2018, pp. 754–757. [Online]. Available: <https://doi.org/10.1109/WI.2018.00010>
- [11] A. J. Smola and B. Schölkopf, “A Tutorial on Support Vector Regression,” *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, 2004.
- [12] D. Ruta, L. Cen, and Q. H. Vu, “Greedy incremental support vector regression,” in *Proceedings of the 2019 Federated Conference on Computer Science and Information Systems, FedCSIS 2019, Leipzig, Germany, September 1-4, 2019.*, M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds., 2019.
- [13] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2006.

# Greedy Incremental Support Vector Regression

Dymitr Ruta  
EBTIC, Khalifa University, UAE  
dymitr.ruta@ku.ac.ae

Ling Cen  
EBTIC, Khalifa University, UAE  
cen.ling@ku.ac.ae

Quang Hieu Vu  
Zalora, Singapore  
quanghieu.vu@zalora.com

**Abstract**—Support Vector Regression (SVR) is a powerful supervised machine learning model especially well suited to the normalized or binarized data. However, its quadratic complexity in the number of training examples eliminates it from training on large datasets, especially high dimensional with frequent retraining requirement. We propose a simple two-stage greedy selection of training data for SVR to maximize its validation set accuracy at the minimum number of training examples and illustrate the performance of such strategy in the context of Clash Royale Challenge 2019, concerned with efficient decks’ win rate prediction. Hundreds of thousands of labelled data examples were reduced to hundreds, optimized SVR was trained on to maximize the validation  $R^2$  score. The proposed model scored the first place in the Cash Royale 2019 challenge, outperforming over hundred of competitive teams from around the world.

**Index Terms**—Support vector regression, greedy backward-forward search, data editing, hyperparameters optimization

## I. INTRODUCTION

Support Vector Machine (SVM) is a supervised machine learning (ML) model developed as far back as in 1963 [1] on the basis of Vapnik-Chervonenkis computational theory of learning [2]. Its introduction brought a breakthrough in back then emerging machine learning domain through the proposition of wide-margin linear separation of classes of data in higher-dimensional input space that otherwise were not separable. Since its original proposal multiple incarnations and advancements have been added, most notably introduction of the non-linear SVM classifier with the kernel trick in [3] and soft margin maximization in [4], [5], shaping SVM to more or less the model we see and use till today.

Support Vector Regression (SVR) extends the original capability of the SVM model into the regression space, while sharing the same model fundamental and properties as SVM does for classification: for instance in margin-maximizing hyper-plane characterization, tolerance of errors etc. With its ground breaking wide-margin generalization capabilities SVM as well as SVR dominated the ML field for decades demonstrating significant improvements in supervised learning problems across many application areas: [1]-[7]

In the face of exponential growth of data in terms of its variety, dimensionality and size, we observe today, SVM (SVR) quadratic complexity in the number of training examples, practically eliminates it from direct applications on large datasets starting from hundreds of thousands of data points, especially if frequent retraining is required [7], [8]. High cost involved in computing large number of support vectors in SVR training process is a critical drawback compared to simpler

supervised ML models, which although unable to demonstrate such generalization ingenuity, are simply able to complete in a reasonable time: [9], [10], [11].

Many SVM (SVR) model efficiency improvements have been proposed recently in an attempt to re-enable the model for the big data world: from simplifications like elimination of linearly dependent support vectors [12], through selective probabilistic examples removal [13], up to support vectors elimination through smoothed separable case approximation [11] or k-mean clustering [8] and more related techniques.

Based on the observation that a vast majority of the SVM (SVR) predictive power comes from fairly small number of key data-structure-capturing examples, an obvious attempt to eliminate huge computational cost of training SVR could be reduced by carefully selecting a small set of the critical training data points. In an attempt to address this challenge we have proposed a simple two-stage greedy search process that returns an ordered list of most predictive data points offering the most predictive SVR model based on incrementally added number of training examples. Combined with automated robust SVR hyper-parameter selection we aspire to achieve a fully automated SVR model construction with a flexible complexity control mechanism. The strength of our model has been thoroughly evaluated in the context of Clash Royale Challenge 2019. This international contest was concerned with construction of the most efficient SVR model to predict win rates of the most popular decks of Clash Royale: a card-based online video game that surpassed 2.5B revenue in the three years since launch. Our parallelizable double-search process was able to reduce the original set of 100000 examples down to 1500 key training data points, which SVR can be trained with near-optimal validation  $R^2$  score. Our method scored the first place in the challenge outperforming more than hundred of participating competitive teams from around the world and offering the gaming platforms an efficient new model for rapid accurate estimation of players win chances to better stimulate their immersion and maintain challenging and immersive engagement.

The remainder of the paper is organized as follows. The Clash Royale Challenge 2019 is described in Section II. The two-stage greedy data selection strategy is presented in Section III, followed with experimental results’ discussion in Section IV and the concluding remarks in Section V.

## II. COMPETITION DESCRIPTION

Clash Royale is a popular video game combining the elements of collectible card game and tower defense genres (<https://clashroyale.com/>). The game involves selecting a deck of 8 playable cards used to attack opponents as well as defend against their cards. The Clash Royale Challenge 2019 is focused on efficient prediction of win rates of the most popular Clash Royale decks in the 1v1 ladder games using support vector regression model. Specifically the intention was to find out whether it is possible to build an efficient win-rate prediction model on a relatively small subset of decks, whose win rates were estimated in the past.

The competition training dataset included 100000 decks comprising exactly 8 cards out of the total of 90 unique possible cards with accompanied win rates computed over 160 million games. The validation set of just 6000 randomly selected decks with win rates was also provided and crucially was extracted from the same period as the true testing set to be used as final evaluation in the competition.

The objective of the competition was to provide 10 subsets of 600,700,...,1500 decks from the training set along with the SVR hyper-parameters of omega, C and gamma, that once trained would result in the highest average  $R^2$  score (Eq. 1) obtained on the testing set unavailable to the competitors. Only preliminary results obtained on the small fraction of the testing set are published on the leaderboard during the competition.

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y}_i)^2} \quad (1)$$

## III. GREEDY 2-STAGE DATA SELECTION FOR SVR

### A. Data preparation

Estimation of future average win rates for every deck was enforced to be done with support vector regression model trained on the bag-of-cards represented decks and their historically computed win rates. Given 90 unique cards the training dataset was transformed to a binary matrix  $X^{[100k \times 90]}$  of 100k (examples) by 90 (card presence indicators), while the output vector  $Y^{[100k \times 1]}$  contained corresponding win rates. Similarly, the validation set  $X_V^{[6000 \times 90]}$  and its corresponding outputs  $Y_V^{[6000 \times 1]}$  were prepared in the same way. Since the validation set was collected from the same period as the unseen testing set it has been decided that the evaluation of any model performance will be obtained using  $R^2$  score computed exclusively on the validation set  $X_V$  against its outputs  $Y_V$ . What it means is that at any point none of the data examples the model is build on will be used to evaluate its performance. Subsequent tests and the leaderboard score feedback positively validated this design choice as a robust generalization feature.

### B. Hyperparameters' setting

The support vector regression model used in the competition used radial basis function (RBF) kernel of the form:

$$G(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (2)$$

In the light of big discrepancies between the training, validation and the leaderboard sets used in the competition we have decided not to optimize  $\gamma$  parameter to the data during training, but rather use the recommended heuristic of setting it to the median distance to the nearest neighbor among randomly selected small subset of the training data.

The constraint to the alpha coefficients, C, was set to the outlier-free estimate of the response  $Y$  standard deviation by setting  $C = IQR(Y)/1.349$ , where  $IQR(Y)$  is the interquartile range of the response variable  $Y$ .

Similarly the  $\epsilon$  parameter is set to 0.1 of the outlier-free estimate of  $Y$ 's standard deviation  $\epsilon = IQR(Y)/13.49$ .

### C. Greedy online backward-forward data selection

SVR training works the fastest with the small number of examples, hence it appears the best option is to ensure the addition of the new data point to the training set maximally improves model's validation performance. Selecting the best new data point requires, however, an exhaustive evaluation of all available remaining data points, which is computationally expensive. A balanced strategy, which we called greedy online backward-forward selection involves a round of sequential additions of any points that improve the current SVR performance followed with rounds of removals that do the same, i.e. improve the current SVR validation performance. To strengthen the reduction side of the process the backward search for removals is repeated until not a single data point's removal improves SVR performance. Such imbalance ensures quicker accumulation of valuable data points and pruning the dataset to the bare minimum, before resuming with the addition, that overall further speeds up SVR training. The advantage of such search is its ability to very quickly find fairly well performing set of training points. The drawback is that it is sequential - hence not parallelizable and lacking the high performance quality of the full exhaustive addition / reduction process. In the competition this search was applied initially to reduce the original set of 100k examples down to 8000 most predictive data points.

### D. Greedy round-exhaustive forward data selection

Greedy round-exhaustive forward data selection follows the simple strategy of adding the best possible data point at each round i.e. adding the point that maximally improves the SVR validation performance. Such search ensures near-optimal performance at the higher computational cost of testing the addition of all other remaining data points before selecting the best at each round. The advantage of such search is also the fact that it is deterministic hence parallelizable at each round. Unlike the greedy online search, it also ensures the important property of incrementally monotonic set performance i.e. its first  $n$  data points are the best  $n$  points of the set. While it is near-intractable to perform such search on the whole set of 100k data points, after reducing it with the fast but sub-optimal greedy online search and together with the parallelized evaluation implementation, it resulted in a relatively fast process of finding incrementally best performing set of 1500

data points. From this set, exploiting the above-mentioned property of incremental performance monotonicity, choosing the best subsets of 600,700,...,1500 was readily given by taking the incrementally growing chunk of the data. The backward side of the greedy backward-forward search was abandoned for this search simply due to its much higher computational cost and relatively low effectiveness since high quality forward search left very little improvement capability for the backward search at the too high computational cost.

#### E. Fine-tuning for further generalization improvements

Despite model's leading leaderboard score, further attempts have been made to further improve its generalization abilities encouraged by still rather big  $R^2$  score discrepancies obtained for training, validation and leaderboard sets. Beside already mentioned robust data-dependent hyper-parameters setting, significant improvement has been also achieved through injecting a little bit of the training set into the validation set such that the validation set gained extra 4000 data points and now amounted to 10000 points in total. The added data have been naturally removed from the training set to avoid training and validating on the same data points. Injection of the data chunk from different period improved validation set diversity and boosted its representativeness, which was reflected in a slight improvement of the leaderboard  $R^2$  score by about 0.01. The increase in the evaluation cost on the larger validation set was to a degree offset by selection from the smaller training set. The composition balance between the training and validation set sizes in the extended validation set was guided by an intuition but certainly further research on optimality of this balance could be conducted with likely further improvements.

### IV. EXPERIMENTAL RESULTS

The above described 2-stage data selection process has been executed on the standalone PC/laptop. The faster greedy online b-f selection has been executed on average performance laptop since it is not parallelizable and yielded fairly quickly the results in a form of about 8000 preselected data points. Throughout this fast search various fine-tuning and generalization boosting strategies in the section above have been tested that led to the chosen automated setting of the SVR hyperparameters and blending the validation set with a small chunk (4000 points) of the training set. Then the greedy round-exhaustive forward search has been executed on the pre-selected 8000 data points to select incrementally near-optimal set of best 1500 points. It has been executed on the standalone DELL PC with 20-cores Xeon processor and the 20-workers *parfor* parallelization utilized to train and evaluate SVR models in each round of data addition. With such setup the execution was also relatively fast and most importantly yielded intermediate results that were mixed with simple complementary selection that yielded incremental score progress on the leaderboard, reassuring the generalization validity of the strategy. The validation set  $R^2$  score obtained on the subset of preselected 8000 points reached in excess of

Table I  
TIMELINE OF MODEL PERFORMANCE IMPROVEMENTS

Component	online	exhaustive	hyperparameters	validation mix
$R^2$ score	0.237	0.258	0.266	0.274

0.6, while the validation scores obtained for the submission-ready 10 solutions of 600, 700, ..., 1500 were in the range of 0.4–0.5. The final leaderboard score of the best solution was almost 0.275 and was the top score among over 100 teams submissions. Although a huge model overfitting has been observed - evident in a form of big differences between the validation set and leaderboard set scores, the consistency and monotonicity of the score improvements achieved throughout submission of the intermediate search results reassured the strategy validity and allow to expect good results.

Based on the feedback from the leaderboard during the competition, Table I reflects the incremental improvements of the  $R^2$  score of the proposed model with gradually added component features throughout the contest duration.

### V. CONCLUSIONS

We have proposed a simple yet robust 2-stage greedy search strategy for selecting a small subset of the incrementally most predictive data points tested with SVR model deployed to learn decks' win rates within Cash Royale Challenge 2019. With the 1<sup>st</sup> place scored by our model we have demonstrated an extreme efficiency of the proposed data editing strategy, which relatively quickly squeezed out the winning accuracy out of only essential 1% of the original 100k dataset, SVR model would otherwise be completely intractable to train on.

### REFERENCES

- [1] V. Vapnik and A. Lerner, "Pattern Recognition Using Generalized Portrait Method", *Automation and Remote Control* 24:774–780, 1963.
- [2] V. Vapnik and A. Chervonenkis, "A Note on One Class of Perceptrons", *Automation and Remote Control* 25, 1964.
- [3] B. Boser, I. Guyon, and V. Vapnik, "A training algorithm for optimal margin classifiers," *Proc. Fifth Annual Workshop of Computational Learning Theory*, vol. 5, pp. 144–152, Pittsburgh, 1992.
- [4] V. Vapnik, "The Nature of Stat. Learning Theory", Springer, NY, 1995.
- [5] C. Cortes and V. Vapnik, "Support-Vector Networks", *Machine learning* 20(3):273-297, 1995.
- [6] V. Vapnik, S. Golowich and A. Smola, "Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing," in M. Mozer, M. Jordan, and T. Petsche (eds.), *Neural Information Processing Systems*, vol. 9, MIT Press, Cambridge, MA., 1997.
- [7] A. Smola, and B. Schölkopf, "A Tutorial on Support Vector Regression," *Statistics and computing*, vol. 14, pp. 199-222, 2003.
- [8] X. Xia, M. Lyu, T. Lok, G. Huang, "Methods of Decreasing the Number of Support Vectors via k-Mean Clustering," *Proc. Int. Conf. Intelligent Computing*, pp. 717-726, 2005.
- [9] C. Burges, "Simplified support vector decision rules," *Proc. 13th Int. Conf. Mach. Learning*, pp. 71-77, 1996.
- [10] E. Osuna and F. Girosi, "Reducing the run-time complexity of support vector machines," *Int. Conf. Pattern Recognition*, Australia, 1998.
- [11] D. Geebelen, J. Suykens, J. Vandewalle, "Reducing the number of support vectors of SVM classifiers using the smoothed separable case approximation", *IEEE Trans Neural Net Learn Sys.* 23(4):682-688, 2012.
- [12] T. Downs, K. Gates, and A. Masters, "Exact simplification of support vector solutions", *Machine Learning Research* 1:293-297, 2001.
- [13] G. Bakir, J. Weston, and L. Bottou, "Breaking SVM complexity with cross-training," *Advances in Neural Information Processing Systems*, vol. 17, pp. 81-88, 2005.



# Training Subset Selection for Support Vector Regression

Cenru Liu

Ngee Ann Polytechnic, Singapore  
liucenru@gmail.com

Jiahao Cen

Nanyang Polytechnic, Singapore  
cenjiahao456@gmail.com

**Abstract**—As more and more data are available, training a machine learning model can be extremely intractable, especially for complex models like Support Vector Regression (SVR) training of which requires solving a large quadratic programming optimization problem. Selecting a small data subset that can effectively represent the characteristic features of training data and preserve their distribution is an efficient way to solve this problem. This paper proposes a systematic approach to select the best representative data for SVR training. The distributions of both predictor and response variables are preserved in the selected subset via a 2-layer data clustering strategy. A 2-layer step-wise greedy algorithm is introduced to select best data points for constructing a reduced training set. The proposed method has been applied for predicting deck’s win rates in the Clash Royale Challenge, in which 10 subsets containing hundreds of data examples were selected from 100k for training 10 SVR models to maximize their prediction performance evaluated using R-squared metric. Our final submission having a  $R^2$  score of 0.225682 won the 3<sup>rd</sup> place among over 1200 solutions submitted by 115 teams.

**Index Terms**—Clash Royal, Support Vector Regression (SVR), R-squared metric ( $R^2$ ), Radial Basis Function kernel (RBF), k-means clustering

## I. INTRODUCTION

NOWADAYS with the growth of the Internet of Things (IoT), 2.5 quintillion bytes of data are produced every day at our current speed [1]. As 2 sides of a coin, a large amount of available data help to build complex and robust machine learning models, while data processing and model training can be rather intractable. Among all data collected, some of them are irrelevant to targets, inter-dependent, and noisy with outliers, leading to inefficient or even intractable training procedure, and more seriously, poor generalization capability.

Support Vector machine (SVM), developed at AT & T Bell Laboratories by Vladimir Vapnik and his co-workers [2], [3], [4], [5], [6], [7] based on the statistical learning theory (or VC theory) [8], [9], [10]. The SVM has shown competitive generalization over many existing machine learning models in various fields, e.g. optical character recognition (OCR), object recognition, time series prediction, etc. [6], [11], [12], [13], [14], as well as in regression, denoted as Support Vector Regression (SVR) [15], [16], [17], [18]. As we know, training a SVR model needs to solve a large quadratic programming optimization problem, which becomes computation intractable on large datasets.

To overcome this disadvantage, it is useful to identify a representative and discriminative data subset from full training data, which is the intention of the Clash Royale Challenge 2019. Clash Royale is a popular video game which combines elements of collectible card game and tower defense genres. In the game, players build decks having 8 cards representing playable troops, buildings, and spells to attack opponent’s towers and defend against their cards. Wining a game is highly dependent on decks. The task of the challenge is to select small data subsets from a large training dataset, on which SVR models can be trained to predict win rates of decks.

To address this problem, a systematic approach is proposed in this paper. The major advantages of our proposed method can be summarized as follows:

- 1) Selecting data points on the clustered space of response variables helps to preserve response distribution, allow parallel implementation, and reduce computational cost.
- 2) Selecting data points from cluster centers of predictor variables can largely speed up search procedure by removing most of training examples from the selection candidates pool, meanwhile reserving predictors’ distribution and their characteristic features.
- 3) Although no guarantee of global optimality, the systematic approach can deterministically find near-optimal solutions.

By using our method in the challenge, 10 subsets containing only hundreds of examples were selected from 100k data points, on which 10 SVR models were trained to predict win rates of decks. The average R-squared metric of the 10 models on unknown testing data is 0.225682, wining 3<sup>rd</sup> place among over 1200 solutions submitted by 115 teams.

This paper is organized as follows. The challenge is described in Section II. The details of the proposed method are presented in Section III. Section IV discusses the experiment results. Conclusions are given in Section V.

## II. CLASH ROYALE CHALLENGE

### A. Challenge task

The intention of the Clash Royale Challenge is to find a small subset from a large training dataset, on which a SVR model with Radial basis function (RBF) kernel can be efficiently trained for predicting win rates of decks. Specifically, competition participants are required to submit 10 subsets of

decks, including 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, and 1500 decks, respectively, each of which allows training an efficient SVR based win rate prediction model, and the hyper-parameters of the SVR trained on the these subsets, i.e.  $\epsilon$ ,  $C$ , and  $\gamma$ .

### B. Database

The data used in the challenge are divided into training, validation, and testing sets. The training data consist of 100k Clash Royale decks that were most commonly used by players during 3 consecutive league seasons in 1v1 ladder games. The decks in the validation and testing data were popular during the three next game seasons after the training data period. The validation dataset consists of 6k decks, which was provided to competitors for self-evaluation of their solutions, while the test set was not revealed to participants. The win rates of decks were also provided in the training and validation datasets. Since the decks in the 2 sets were collected from different game seasons, the same decks in different sets may have different win rates.

### C. Solution evaluation

The quality of solutions is assessed using prediction performance measured in the R-squared metric of the models trained on the indicated subsets and the associated hyper-parameters. The R-squared metric is defined as

$$R^2 = 1 - \frac{RSS}{TSS}, \quad (1)$$

where  $RSS$  is the residual sum of squares and  $TSS$  is the total sum of squares, which can be expressed as

$$RSS = \sum_i (y_i - f_i)^2, \quad (2)$$

and

$$TSS = \sum_i (y_i - \frac{1}{N} \sum_i y_i)^2, \quad (3)$$

where  $y_i$  and  $f_i$  are the ground truth label of the  $i^{th}$  data example and its prediction, respectively, and  $N$  is the number of data records in the dataset. The score of a solution is the average  $R^2$  metric of the 10 SVR models.

Leaderboard scores were provided in the preliminary stage of the challenge, which were calculated based on a small subset of the testing data fixed to all participants. The final scores of the 2 best solutions submitted by a competitor evaluated on the full testing set were provided at the end of the challenge.

## III. METHOD FOR SUBSET SELECTION

### A. Method overview

A systemic method is proposed to select a small subset of data for training an efficient SVR model, which consists of 5 parts concluded as follows, as shown in Fig. 1.

- 1) Dividing training data into  $k_y$  groups according to the response variable, denoted as  $y$ , e.g. win rates in the challenge.

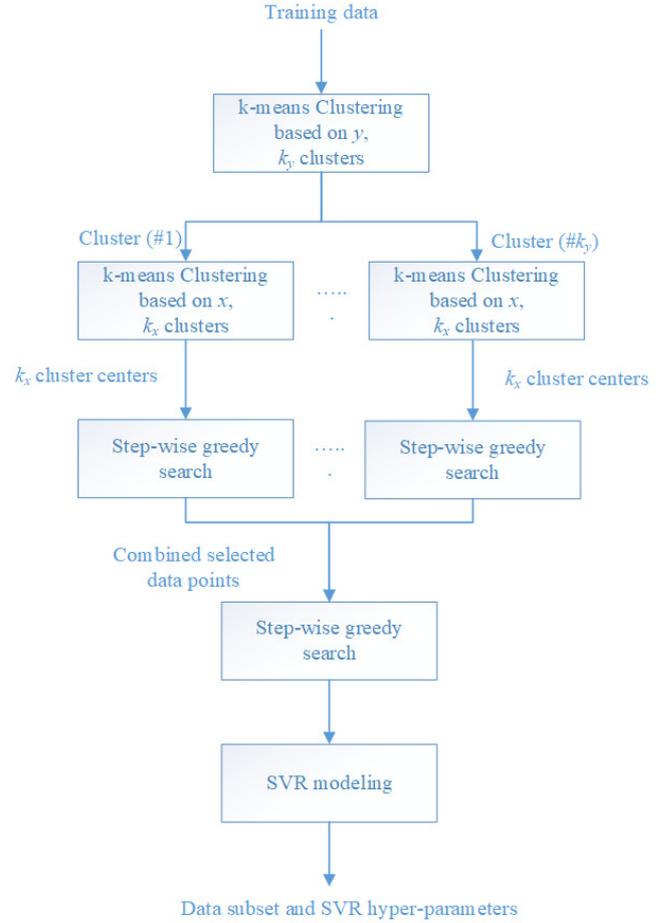


Figure 1. Flowchart of the proposed method.

- 2) Dividing each of  $k_y$  groups into  $k_x$  clusters according to predictor variables, denoted as  $x$ , e.g. decks in the challenge, and constructing  $k_y$  sets of cluster centers.
- 3) Selecting a specific number of data points individually from each of  $k_y$  center-sets by step-wise greedy search. The number is dependent on the sizes of the full dataset, center-set and the subset to be constructed, which will be discussed later. Note that the total number of selected points should be much more than the desired size of the subset.
- 4) Combining all points selected from the  $k_y$  center-sets and selecting exact number of points to construct the required subset by applying again the step-wise greedy algorithm.
- 5) Obtaining the settings of hyper-parameters ( $\epsilon$ ,  $C$ , and  $\gamma$ ) for the SVM model trained on the selected subset.

### B. Data representation

A data example is represented using a binary vector with a length of 90 representing 90 unique cards. Each value in the vector indicates whether or not a card is in the deck, i.e.

- 1- the associated card is used in the deck,

- 0- the associated card is not used in the deck.

The training data containing 100k examples are represented using a matrix with a dimension of  $100000 \times 90$  and the validation data having 6000 examples are represented using a matrix with a dimension of  $6000 \times 90$ . The response variable of the training data, i.e. win rates, is represented using a vector with a length of 100000, and similarly, the win rates of the validation set are represented using a vector with a length of 6000.

It has been mentioned in Section II that the decks in training and validation sets were extracted from different game seasons. Although the same decks may exist in both sets, their win rates are likely different because the game evolves in time, players adapt to new strategies, and the balance of individual cards and their popularity changes slightly from one season to another. Removing the training examples having the same decks as the validation set but with different win rates can avoid uncertainty of such gap, which, however, cannot yield significant improvement on prediction accuracy. This indicates, from a certain of view, the robustness of our selection method.

### C. Two-layer clustering analysis

Clustering analysis is firstly applied to guild data selection. Specifically, a 2-layer clustering strategy inspired by the work presented in [19] is employed to divide training data into groups, as illustrated in Fig. 1. In our method, data clustering is performed by using the K-means clustering algorithm that is a classical and popular unsupervised machine learning algorithm [20]. The aim of clustering analysis here is to preserve the distribution of the full training dataset and reflect their characteristic features in a reduced dataset.

Clustering analysis is performed independent on predictor and response variables, e.g. decks and win rates in the challenge.

- 1) The training dataset is firstly separated into  $k_y$  clusters according to the response variable. The value of  $k_y$  can be set empirically based on the distribution of  $y$ , e.g.  $k_y = 2$  in win rate prediction. In this way, the distribution of  $y$  can be preserved, and meanwhile the subsequent steps can be implemented in parallel.
- 2) Each of  $y_k$  groups are then further divided into  $k_x$  clusters according to the predictor variables. The value of  $k_x$  is empirically determined according to the distribution of  $x$  as well as the sizes of training dataset and the subset to be selected.

We can finally obtain  $k_y$  groups, each having  $k_x$  cluster centers, via the 2-level clustering strategy. Similarly, the validation dataset can be divided into groups using the same cluster centers as the training data.

### D. Two-layer step-wise greedy search

The data subset is selected to feed to SVR training to maximize the prediction performance of the model via a 2-layer step-wise greedy search strategy .

- 1) First, a specific number of data points are independently selected from each of  $k_y$  center-sets by step-wise greedy search that follows below procedure, where  $X$  denotes the full training set containing  $N$  data points,  $S$  represents the subset to be built and  $R(S)$  is its  $R^2$  score.

- Step 1. The search procedure starts with a full training set of  $X$  and an empty subset of  $S$ .
- Step 2. Adding the data point, denoted as  $p$ , selected from  $X$  to  $S$ , which gives the highest score among all points in  $X$ .
- Step 3. Removing the  $p^{th}$  point from  $X$ , and  $N = N - 1$ .
- Step 4. Going to Step 2 until  $S$  is fully filled.

The score of a SVR model is the R-squared metric given in (1) calculated on the validation dataset.

Let  $N_i$  be the number of data points selected from the  $i^{th}$  center-set, which is set as:

$$N_i = N_{all} \times \left( \frac{c_{ti}}{N_t} + \frac{c_{vi}}{N_v} \right) / 2, \quad (4)$$

for  $i \in [1, 2, \dots, k_y]$ , where

- $N_{all}$  is the approximate total number of data points to be selected from all of  $k_y$  clusters, which can be empirically set to be twice as the desired size of the data subset under selection;
- $c_{ti}$  and  $c_{vi}$  are the sizes of the  $i^{th}$  center-sets of the training and validation sets, respectively;
- $N_t$  and  $N_v$  are the sizes of the full training and validation sets, respectively.

- 2) After the data points are selected from each of  $k_y$  center-sets, they are combined to construct a bigger set, on which the step-wise greedy search is applied again to select best data points based on the same selection criteria as the first layer of greedy search.

### E. SVR hyper-parameters

The hyper-parameters of the non-linear SVR model with a Gaussian radial basis function kernel, including  $\epsilon$ ,  $C$ , and  $\gamma$ , are optimized for each selected subset using a heuristic grid search with a range around the seeds and a grid of 0.00001. The seeds of the hyper-parameters are set as follows.

- 1)  $\epsilon$  in the  $\epsilon$ -insensitive loss function controls the smoothness of the SVR model and the number of support vectors, which can largely affect model complexity and its generalization capability.  $\epsilon$  is set to be an estimate of a tenth of the standard deviation using the inter-quartile range of the response variable  $y$ , expressed as:

$$\epsilon = iqr(y) / 13.49, \quad (5)$$

where  $iqr(y)$  is the inter-quartile range of  $y$ .

- 2) The parameter  $C$  controls the trade off between training error and model complexity, i.e. margin maximization, e.g.  $C = \infty$  yielding a hard margin SVR model. In our method,  $C$  is set to be an estimate of the standard deviation of the response variable, expressed as:

$$C = iqr(y) / 1.349. \quad (6)$$

- 3)  $\gamma$  is a free parameter used in the radial kernel. The radial basis function kernel, or RBF kernel on two samples  $x_i$  and  $x_j$  is defined as

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2). \quad (7)$$

The value of  $\gamma$  is optimized by the heuristic procedure using sub-sampling [21].

#### IV. EXPERIMENT RESULTS

The numbers of clusters in the 2-layer clustering analysis were set to be:

$$k_y = 2, \quad (8)$$

i.e. the data were divided into 2 clusters according to win rates, and

$$k_x = 5000, \quad (9)$$

i.e. the data in each of the 2 groups were divided into 5000 clusters. The full training dataset containing 100k examples were reduced into 10k cluster centers from 2-layer clustering analysis, among which 10 relative small subsets containing the required numbers of data examples were selected by using the 2-layer step-wise greedy search strategy.

The best solution that we submitted to the competition as the final solution has a preliminary R-squared metric of 0.2352 evaluated on a subset of testing data and a final score of 0.225682 evaluated on the full testing set, which was scored the 3<sup>rd</sup> place in the challenge among over 1200 solutions submitted by 115 teams.

Although the current version of the proposed method was designed to select a best data subset for SVR model training, our method can be easily extended for other machine learning methods without many modifications. The search procedure followed in our method adding data points in a recursive way cannot guarantee global-optimal performance. Improvement can be expected with suitable implementation of global search.

#### V. CONCLUSIONS

It is useful to select a subset from full labeled data for efficiently training machine learning models, in order to maximize prediction performance at a small number of data examples. This cannot only reduce computational cost but also lead to better generalization capability. To address this, a systematic approach is proposed for data selection, the performance of which has been shown in the Clash Royale Challenge, in which 100k data points were reduced to 600-1500 inputted to train Support Vector Regression (SVR) based win rate prediction models, winning the 3<sup>rd</sup> place in the challenge. This method, although developed for data selection in SVR training, can be easily modified for other machine learning methods. Future work will also improve the search procedure by introducing global optimization methods like evolutionary algorithms.

#### REFERENCES

- [1] B. Marr, "How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read," <https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#ae77e2560ba9>, 2018.
- [2] B.E. Boser, I.M. Guyon, V. Vapnik, "A training algorithm for optimal margin classifiers," *Proceedings of the Annual Conference on Computational Learning Theory, ACM*, pp. 144–152, Pittsburgh, PA 1992.
- [3] I. Guyon, B. Boser, and V. Vapnik, "Automatic capacity tuning of very large VC-dimension classifiers," *Advances in Neural Information Processing Systems 5*, pp. 147–155, Morgan Kaufmann Publishers, 1993.
- [4] C. Cortes, and V. Vapnik, Support vector networks, Machine Learning, vol. 20, pp. 273–297, 1995.
- [5] B. Schölkopf, C. Burges, and V. Vapnik, "Extracting support data for a given task," *Proceedings of First International Conference on Knowledge Discovery and Data Mining, AAAI Press*, 1995.
- [6] B. Schölkopf, C. Burges, and V. Vapnik, "Incorporating invariances in support vector learning machines," *Artificial Neural Networks, Springer Lecture Notes in Computer Science*, Vol. 1112, pp. 47–52, Berlin, 1996.
- [7] V. Vapnik, S. Golowich and A. Smola, "Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing," in M. Mozer, M. Jordan, and T. Petsche (eds.), *Neural Information Processing Systems*, vol. 9, MIT Press, Cambridge, MA., 1997.
- [8] V. Vapnik and A. Chervonenkis, "Theory of Pattern Recognition" (in Russian), Nauka, 1974.
- [9] V. Vapnik, "Estimation of dependences based on empirical data," Springer Verlag.
- [10] V. Vapnik, "The Nature of Statistical Learning Theory," Springer, New York.
- [11] B. Schölkopf, P. Simard, A. Smola, and V. Vapnik, "Prior knowledge in support vector kernels," In: M.I. Jordan, M.J. Kearns, and S.A. Solla (Eds.), *Advances in Neural Information Processing Systems 10*, MIT Press, Cambridge, MA, pp. 640–646, 1998.
- [12] V. Blanz, B. Schölkopf, H. Bülthoff, C. Burges, V. Vapnik, and T. Vetter, "Comparison of view-based object recognition algorithms using realistic 3D models," *Artificial Neural Networks, Springer Lecture Notes in Computer Science*, vol. 1112, pp. 251–256, Berlin, 1996.
- [13] B. Schölkopf, K. Sung, C. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. Vapnik, "Comparing support vector machines with Gaussian kernels to radial basis function classifiers," *IEEE Transactions on Signal Processing*, vol. 45, pp. 2758–2765, 1997.
- [14] K.R. Muller, A. Smola, G. Ratsch, B. Schölkopf, J. Kohlmorgen, and V. Vapnik, "Predicting time series with support vector machines," *Artificial Neural Networks, Springer Lecture Notes in Computer Science*, vol. 1327, pp. 999–1004, Berlin, 1997.
- [15] H. Drucker, C.J.C. Burges, L. Kaufman, A. Smola, and V. Vapnik, "Support vector regression machines," *Advances in Neural Information Processing Systems 9*, pp. 155–161, MIT Press, Cambridge, MA, 1997.
- [16] M. Stitson, A. Gammerman, V. Vapnik, V. Vovk, C. Watkins, and J. Weston, "Support vector regression with ANOVA decomposition kernels," *Advances in Kernel Methods—Support Vector Learning*, MIT Press Cambridge, MA, pp. 285–292, 1999.
- [17] A. Smola, and B. Schölkopf, "A Tutorial on Support Vector Regression," *STATISTICS AND COMPUTING*, vol. 14, pp. 199–222, 2003.
- [18] D. Basak, S. Pal, and D. Patranabis, "Support Vector Regression," *Neural Information Processing – Letters and Reviews*, vol. 11, Non. 10, pp. 203–224, October 2007.
- [19] X. Xia, M. Lyu, T. Lok, G. Huang, "Methods of Decreasing the Number of Support Vectors via k-Mean Clustering," *Proc. International Conference on Intelligent Computing, Lecture Notes in Computer Science book series (LNCS)*, vol. 3644 pp. 717–726, 2005.
- [20] J. Hartigan, and M. Wong, "Algorithm AS 136: A k-Means Clustering Algorithm," *Journal of the Royal Statistical Society, Series C*, vol. 28, no. 1, pp. 100–108, 1979.
- [21] fitrsvm: Fit a support vector machine regression mode, <https://www.mathworks.com/help/stats/fitrsvm.html>.

# Efficient Support Vector Regression with Reduced Training Data

Ling Cen

EBTIC, Khalifa University, UAE  
cen.ling@kustar.ac.ae

Quang Hieu Vu

Zalora, Singapore  
quanghieu.vu@zalora.com

Dymitr Ruta

EBTIC, Khalifa University, UAE  
dymitr.ruta@kustar.ac.ae

**Abstract**—Support Vector Regression (SVR) as a supervised machine learning algorithm have gained popularity in various fields. However, the quadratic complexity of the SVR in the number of training examples prevents it from many practical applications with large training datasets. This paper aims to explore efficient ways that maximize prediction accuracy of the SVR at the minimum number of training examples. For this purpose, a clustered greedy strategy and a Genetic Algorithm (GA) based approach are proposed for optimal subset selection. The performance of the developed methods has been illustrated in the context of Clash Royale Challenge 2019, concerned with decks' win rate prediction. The training dataset with 100,000 examples were reduced to hundreds, which were fed to SVR training to maximize model prediction performance measured in validation  $R^2$  score. Our approach achieved the second highest score among over hundred participating teams in this challenge.

**Index Terms**—Support Vector Regression (SVR), K-means clustering, greedy search, R-squared metric, Clash Royale

context of Clash Royale Challenge 2019 with an aim to build an efficient win-rate prediction model on a relatively small subset of decks. The 100,000 labelled data examples in the training dataset were reduced to hundreds, over which a SVR model can be trained with near-maximal validation  $R^2$  score. Our method achieved the second highest score among over hundred participating teams. In addition, a Genetic Algorithm (GA) based approach is also proposed for subset selection to explore global search in training data reduction.

The remainder of the paper is organized as follows. The Clash Royale Challenge 2019 is described in Section II. The clustered greedy selection strategy is elaborated in Section III, followed with the GA based selection approach in Section IV. The experiment results are discussed in Section V. Finally, concluding remarks are given in Section VI.

## I. INTRODUCTION

Support Vector Regression (SVR) shares the same set of properties as Support Vector Machine (SVM) does for classification. Examples include tolerating some errors, characterizing hyper-plane that maximizes the margin, etc. Because of these good properties, during the past decades, SVR as well as SVM have attracted increasing interest and successfully solved supervised machine learning problems in various fields [1], [2], [3]. Its quadratic complexity in the number of training examples, however, eliminates the SVR from training on large datasets, especially if frequent retraining is required [4], [5]. High computational cost associated with the large number of support vectors is a critical drawbacks in comparison with other supervised machine learning algorithms [6], [7], [8].

To improve model efficiency, some approaches for model simplification have been proposed in the literature, e.g. eliminating support vectors linearly dependent on the other support vectors [9], selectively removing examples from training data using probabilistic estimates related to editing algorithms [10], reducing the number of support vectors using smoothed separable case approximation [8] or k-mean clustering [5], etc.

One efficient way for fast SVR training is to maximize its prediction accuracy at the minimum number of training examples. To address this challenge, a multi-step clustered greedy strategy is proposed for selecting a small data subset fed to SVR training fitted with automated robust hyper-parameter selection. Its performance has been illustrated in the

## II. COMPETITION DESCRIPTION

Clash Royale is a popular video game, where players build decks consisting of 8 cards representing playable troops, buildings, and spells to attack opponent's towers and defend against their cards. Building good decks is, therefore, critical to win the game. The intention of the challenge is to find out whether it is possible to build an efficient win-rate prediction model on a relatively small subset of decks, whose win rates were estimated in the past.

The competition training dataset includes 100,000 decks comprising 8 cards out of the total of 90 unique possible cards, which were most commonly used by players during 3 consecutive league seasons in 1v1 ladder games, with accompanied win-rates computed over 160m games. The validation set contains 6000 randomly selected decks with their corresponding win-rates, which was extracted from the 3 next game seasons after the training data period. The testing data extracted from the same period as the validation set, which were unrevealed to participants, were used to evaluate the solutions submitted to the competition,

The task of the competition was to select 10 subsets from the 100,000 training decks, on which 10 efficient SVR models can be trained with best performance of win-rate prediction. Besides the 10 subsets, the hyper-parameter values of the SVR models with radial kernels, including  $\epsilon$ ,  $C$ , and  $\gamma$ , were required together.

The Performance of a SVR model is assessed by the  $R^2$  metric of the model, which is defined as

$$R^2 = 1 - \frac{\sum_i (y_i - p_i)^2}{\sum_i (y_i - \frac{1}{N} \sum_i y_i)^2}, \quad (1)$$

where  $y_i$  and  $p_i$  are the true and predicted values of the win-rate of the  $i^{th}$  data point, respectively, and  $N$  is the size of the testing dataset. The score of a solution is the average of the  $R^2$  scores of the 10 SVR models.

The facility to score derived model solutions on a part of the testing set was provided via the web-based KnowledgePit platform. Although the submission had to be evaluated for the whole testing set, the feedback in a form of the  $R^2$  score was received based on a small subset of the testing examples, fixed for the competitors in the preliminary stage

### III. CLUSTERED GREEDY SELECTION STRATEGY

The clustered greedy strategy has been developed for selecting optimal training subsets, which consists of 4 steps, i.e. k-means clustering, forward greedy search, sequence optimization, and fine-tuning process. The implementation details of the method will be elaborated in this section.

#### A. Data preparation

Estimation of future average win-rates for every deck are enforced to be done with the SVR model trained on the bag-of-cards represented decks and their historically computed win-rates. Given 90 unique cards the training dataset is transformed to a binary matrix with a dimension of  $100k \times 90$  representing 100k (examples) by 90 (card presence indicators), while the output vector with a dimension of  $100k \times 1$  contains corresponding win-rates. Similarly, the validation set ( $6000 \times 90$ ) and its corresponding outputs ( $6000 \times 1$ ) are prepared in the same way.

There is a big gap between the validation  $R^2$  values of our submission-ready solutions and the leaderboard scores, e.g. the former is in the range of 0.4-0.5 while the latter is in 0.2-0.25. To avoid over-fitting and achieve robust models, the validation dataset are extended by combining the original validation set and training data in 4 ways, denoted as E1, E2, E3, and E4, which are:

- E1: 6000 data examples in the original validation dataset;
- E2: 6000 data examples in the original validation dataset and 6000 examples having the largest number of games in the training dataset;
- E3: all data examples in the training dataset, and 16 copies of the original validation dataset for balanced involvement of training and evaluation dataset;
- E4: removing the training points having the same decks as those in the validation set from E3 due to the big discrepancies between the two sets.

The performance of SVR models obtained during search will be evaluated on one of the 4 validation sets.

#### B. Hyper-parameters of SVR

The hyper-parameters of the SVR models with radial basis function (RBF) kernel, including  $\epsilon$ ,  $C$ , and  $\gamma$ , are achieved in below ways:

- $C$ , the constraint to the alpha coefficients, is set as  $C = iqr(Y)/1.349$ , where  $iqr(Y)$  is the inter-quartile range of the response variable,  $Y$ .
- $\epsilon$  is set to be an estimate of 0.1 of  $Y$ 's standard deviation, i.e.  $\epsilon = iqr(Y)/13.49$ .
- $\gamma$  is selected using the heuristic procedure internally implemented in MATLAB.

#### C. k-means clustering

The idea here is to constitute a subset with the data points distributed in the full space of training data. To achieve this, the data are firstly divided into  $k$  groups by k-means clustering. A subset is composed by selecting data points equally from each of the  $k$  clusters. Smaller  $k$  leads to high computational cost and possibly over-fitting caused by concentrated distribution of the selected data, while bigger  $k$  may overlook unique distribution of the training data. We have made a comparison on different values of  $k$ , e.g. 20, 50, and 100, from which it can be seen that  $k = 50$  gives the best results.

#### D. Forward greedy search (FGS)

After dividing the training data into  $k$  clusters, a forward greedy algorithm is applied to select the best subset, which follows a simple strategy of adding the best possible data point from one cluster at each time. After a round is completed, in which  $k$  points have been added respectively from  $k$  clusters, a new round is started if the subset is not fully filled. The flowchart of the search process are shown in Fig. 1.

This search ensures near-optimal performance at the high computational cost of testing the addition of all remaining data points before selecting the best at each search. The advantage of our method is exhaustive evaluation is only performed on data points within a cluster, which, compared to testing all points in the full training dataset, reduces computational cost to  $1/k$ . In addition, such search is deterministic hence it can be implemented in parallel.

Below list compares the regression performance of the SVR models trained over the subsets selected with different values of  $k$  and using different validation sets for performance evaluation of any model yielded in search:

- $R^2 = 0.2158$ ,  $k = 100$ , validation set: E1,
- $R^2 = 0.2277$ ,  $k = 50$ , validation set: E1,
- $R^2 = 0.2566$ ,  $k = 20$ , validation set: E2,
- $R^2 = 0.2593$ ,  $k = 50$ , validation set: E2,

where  $R^2$  is the leaderboard score received in the preliminary stage.

#### E. Sequence optimization (SO)

The greedy search that chooses what appears to be the optimal immediate choice at each time cannot ensure global optimal performance since the current best point may not lead to global best path. To improvement this, after 1500 training

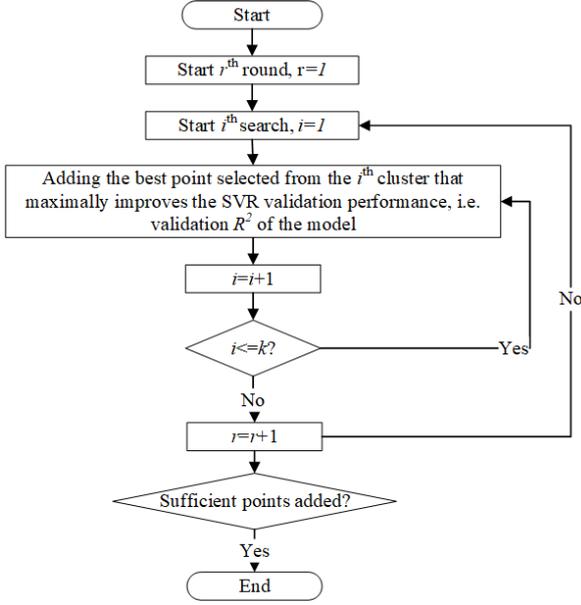


Figure 1. Flowchart of forward greedy search, where  $i$  and  $r$  denote the indices of a search and a round, respectively, i.e. the  $i^{\text{th}}$  search is to find the best point from all available remaining data in the  $i^{\text{th}}$  cluster, and a round is to find  $k$  points in  $k$  clusters respectively.

data points are selected, the sequence of the selected data points are re-arranged by starting a new round of forward greedy search within the 1500 points. Searching within a compressed set, model evaluation can be performed on E3 or E4 validation set at a much lower computational cost than exhaustive evaluation of all available points in a full clustered set. The prediction performance in the steps of FGS and SO are compared below ( $k = 50$ ):

- $R^2 = 0.2277$  in FGS with a validation set of E1  $\rightarrow$   $R^2 = 0.2445$  in SO with a validation set of E3;
- $R^2 = 0.2593$  in FGS with a validation set of E2  $\rightarrow$   $R^2 = 0.2655$  in SO with a validation set of E3;
- $R^2 = 0.2593$  in FGS with a validation set of E2  $\rightarrow$   $R^2 = 0.2702$  in SO with a validation set of E4.

After sequence optimization, the first  $n$  data points in the selected subset are the best  $n$  points of the set. From this set choosing the best 600,700,...,1500 is readily given by taking the incrementally growing chunk of the data.

#### F. Fine-tuning process

The final step of our selection approach is a fine-tuning process, which constitutes a new subset by combining the support vectors of the SVR model from the previous step with the training examples outside  $\epsilon$ -intensive band with smaller deviation between ground truth and corresponding prediction. The improvement, however, is not always quite obvious. The solution with  $R^2 = 0.2702$  can only be improved to  $R^2 = 0.2703$ , while some solutions achieved in previous steps can be improved a little more, e.g. the solution with  $R^2 = 0.2593$  can be improved to  $R^2 = 0.2606$ .

## IV. GA BASED SELECTION APPROACH

In addition to the main method that was presented in the previous section, we also implemented another approach using Genetic Algorithm (GA). In this section, we will present our GA based approach to select training decks.

### A. Population, individual (chromosome), and gene

In GA, at any point of time, there is a population consisting of individuals each of which is a possible solution that includes ten different sets of training decks together with the three required parameters to train an SVM:  $\epsilon$ ,  $C$ , and  $\gamma$ . In other words, an individual in our GA population is a possible solution or submission to the competition.

Given the above definition for an individual in GA, we can see that there are a couple of ways to define a gene in the individual (as an individual is a chromosome that contains a set of genes).

- A possible definition is to consider each training set of deck indices together with the parameters  $\epsilon$ ,  $C$ , and  $\gamma$  as a gene. In this way, we have exactly 10 genes from 10 training sets of decks in each individual. This definition, however, has an issue as the gene is too big to efficiently and effectively perform different variation operations.
- Instead of applying the above definition, the smallest unit of the individual is considered as a gene in which a gene could be a specific  $\epsilon$ ,  $C$ , and  $\gamma$  to train a model with a set of training decks, or even a training deck in this training set. While this definition gives us a finer granulation for the gene, it requires some tricks to support crossover and mutation that we will discuss later to make generation evolve.

### B. First generation

As in a typical approach, the first generation of GA should be generated randomly.

- A random  $\epsilon$  in the range: 0.0 to 1.0
- A random  $C$  in the range: 0.0 to 1000.0
- A random  $\gamma$  in the range: 0.0 to 10.0
- A random set of indices in the range: 1.0 to 100000.0

However, in order to help the GA involve faster, in addition to randomly generated individuals, we also employ few simple approaches to get some seed (good) individuals for the first generation. Note that these approaches are only used to select (possible) better training decks (indices). For  $\epsilon$ ,  $C$ , and  $\gamma$ , we use default values (specifically,  $\epsilon = 0.1$ ,  $C = 1.0$ , and  $\gamma = 1.0/90$ ). Training deck indices were generated for the seed individuals in the following approaches

- Using indices from decks having the highest number of games in the training data.
- Using indices from decks having the highest number of players in the training data.
- Using k-means algorithm to cluster training data into different groups (e.g. 60  $\rightarrow$  150) and again selecting the top-10 indices from each group having the highest number of games or the highest number of players.

### C. Fitness measurement

As each individual in our GA is a possible solution or submission, the straightforward fitness score is the prediction score of the validation data using model trained by parameters and indexed data specified in the individual. In addition to this fitness measurement, another way is to evaluate the model using both training and validation data sets (using different way to give higher weights to the validation data than the training data – as ultimately, we still need to mainly rely on the validation data set). While this validation seems to be better to avoid over-fitting, the trade-off, however, is that it takes significantly more time for the evaluation as the model needs to be evaluated for a much bigger set of data.

### D. Mutation

Given an individual, we first randomly select a set of training decks for the mutation. Then, we will choose to change one of the following components:

- Changing  $\epsilon$  to a random number between 0.0 and 1.0, changing  $C$  to a random number between 0.0 and 1000.0, and changing  $\gamma$  to a random number between 0.0 and 10.0 all with a grid of  $10^{-6}$ .
- Initially, we randomly selected a single training deck to be replaced by another one outside the training set. However, this approach makes the GA extremely slow in progress. Thus, instead of selecting one training deck, to make the GA evolve faster, we chose to randomly select 5% of training decks from the existing indices for replacement.

Note that in each generation, we randomly select 25% of the population to apply mutation for generating new individuals.

### E. Crossover

Given a pair of individuals, we first randomly select a set of training decks for crossover. Then, we choose to perform crossover in the following components:

- Choosing an  $\epsilon$ ,  $C$ ,  $\gamma$  independently from a randomly selected individual.
- We first randomly select a training set of indices having the same size from both individual (e.g. training set of 1000 indices from both sides). Then, from the two training set of indices, we select half of them from each individual. Note that there could be overlapping in the selected indices from both individual, and hence generate less than the number of required indices. In this case, we have two ways to fill the missing indices: to continuously select indices from two individuals to fill or randomly select new indices from outside to fill. In our approach, we randomly use one of the two methods.

Note that in each generation, we randomly select 50% of the population to apply crossover for generating new individuals.

### F. Selection

We follow the traditional approach to select individuals from a generation to the next one. Basically, the probability of an

individual to be selected is proportional to the fitness score it has. It means that the stronger (higher fitness score) of an individual, the higher chance it is being selected to be in the next generation. In our implementation, we choose to maintain a population of 50 individuals in each generation.

## V. EXPERIMENT RESULTS

By applying the proposed methods, the best solution was achieved by clustered greedy selection with below settings:

- $k = 50$  in clustering,
- validation set: E2 in the step of FGS,
- validation set: E4 in the step of SO.

Its leaderboard  $R^2$  score is 0.2593 from forward greedy search and improved to 0.2703 via sequence optimization and fine-tuning. The SO contributes most to score improvement from 0.2593 to 0.2702. The final score evaluated on the full testing dataset is 0.253017. Both the preliminary and final scores of the solution are the 2<sup>nd</sup> highest among over hundred participating teams, showing robustness of the method against over-fitting.

## VI. CONCLUSIONS

This paper explores the possibility of training a Support Vector Regression (SVR) model using a minimal number of training data samples. Two approaches, i.e. clustered greedy strategy, and Genetic Algorithm (GA) based method, are proposed for the selection of data subset fed to SVR training to maximize validation performance. The details of the implementation are elaborated in the paper. The proposed methods successfully selected hundreds of points from 100,000 labeled data samples for efficient SVR training in decks' win-rate prediction and scored 2<sup>nd</sup> place among over hundred participating teams in the Clash Royale Challenge 2019.

## REFERENCES

- [1] B. Boser, I. Guyon, and V. Vapnik, "A training algorithm for optimal margin classifiers," *Proc. Fifth Annual Workshop of Computational Learning Theory*, vol. 5, pp. 144–152, Pittsburgh, 1992.
- [2] V. Vapnik, "The Nature of Statistical Learning Theory," Springer, New York, 1995.
- [3] V. Vapnik, S. Golowich and A. Smola, "Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing," in M. Mozer, M. Jordan, and T. Petsche (eds.), *Neural Information Processing Systems*, vol. 9, MIT Press, Cambridge, MA., 1997.
- [4] A. Smola, and B. Schölkopf, "A Tutorial on Support Vector Regression," *Statistics and computing*, vol. 14, pp. 199–222, 2003.
- [5] X. Xia, M. Lyu, T. Lok, G. Huang, "Methods of Decreasing the Number of Support Vectors via k-Mean Clustering," *Proc. Int. Conf. Intelligent Computing*, pp. 717–726, 2005.
- [6] C. Burges, "Simplified support vector decision rules," *Proc. 13th Int. Conf. Mach. Learning*, pp. 71–77, 1996.
- [7] E. Osuna and F. Girosi, "Reducing the run-time complexity of support vector machines," *Int. Conf. Pattern Recognition*, Australia, 1998.
- [8] D. Geebelen, J. Suykens, J. Vandewalle, "Reducing the number of support vectors of SVM classifiers using the smoothed separable case approximation," *IEEE Trans Neural Netw Learn Syst.*, vol. 23, no. 4, pp. 682–688, 2012.
- [9] T. Downs, K. Gates, and A. Masters, "Exact simplification of support vector solutions," *Journal of Machine Learning Research*, vol. 1, pp. 293–297, 2001.
- [10] G. Bakir, J. Weston, and L. Bottou, "Breaking SVM complexity with cross-training," *Advances in Neural Information Processing Systems*, vol. 17, pp. 81–88, 2005.

# Information granule system induced by a perceptual system

Anna Bryniarska

Opole University of Technology

Institute of Computer Science

ul. Proszkowska 76, 45-758 Opole, Poland

Email: a.bryniarska@po.opole.pl

**Abstract**—Knowledge represented in the semantic network, especially in the Semantic Web, can be expressed in attributive language *AL*. Expressions of this language are interpreted in different theories of information granules: set theory, probability theory, possible data sets in the evidence systems, shadowed sets, fuzzy sets or rough sets. In order to unify the interpretations of expressions for different theories, it is assumed that expressions of the *AL* language can be interpreted in a chosen relational system called a granule system. In this paper, it is proposed to use information granule database and it is also demonstrated that this database can be induced by the measurement system of the adequacy of information retrieval, called a perceptual system. It can simplify previous formal description of the information granule system significantly. This paper also shows some examples of inducing rough and fuzzy granule databases by some perceptual systems.

## I. INTRODUCTION

IT IS intuitively assumed that **conceiving of information**, represented by descriptions of something, is to discern, distinguish, and identify this thing. Conceiving of information about an object is preceded by **the perception** of the description of this object. The perception consists of a degree of compliance between certain information resources about the object and precisely determined knowledge represented in the set of object descriptions called **the thesaurus** [9], [24]. Thus, object perception determines the weight, rank, and importance of object descriptions representing information about this object. This also applies to sources of information about objects, pointing to these objects, called in the computer science **entities**, i.e. such signs of these objects, which are different from their descriptions. Each such reference is called **the information granule** [24], [25] and its instance is called *data* about the object that this information is concerned with. The description of the information granule indicates what this information is about. In the Web, any description, and thus the description of the information granule, has the address of information in the memory of computers connected to this network. This address indicates the sign for human of what the information relates to, including a specific description of the object. These are, for example, natural language expressions describing these objects, data representative about these objects, their image or their sound characteristics. Granules are grouped into granule systems in which **granular calculations** are made, i.e. the information about objects is interpreted. For

well-established knowledge, granules are *data sets*. When, for knowledge representation, incorrect classification of objects is used, i.e. knowledge about them is probable, uncertain, unclear, or vague, then information granules can not be described by abstract data sets. In such situation, to determine the information granules it is proposed to use the following nonstandard formalisms of the set theory [22]: interval analysis [14], fuzzy sets [29], [30], [31], rough sets [15], [16], [17], [18], [19] and shadowed sets [20], [21], [22]. In the papers mentioned above, as well as in other papers on granular calculations, there is a lack of uniform methods of using abstraction in order to interpret the expressions of the attributive language *AL* in the information granule theories.

It should be noted that in computer science, from the beginning of its existence, abstraction has been used to reduce the complexity of the problem and achieve greater transparency [27]. The elementary form of abstraction introduces a distinction between the level of a concrete (instance, instance of data) and its type. With abstraction, the class type of similar concretes can be specified. The lowest level of abstraction is one that does not require skipping (abstaining from) significant differences between objects. The abstraction model is an abstract set of data about objects, described in the terminology of the set theory in the Cantor sense. At present, the sets are understood as such abstractions in which the formal language of the Zermelo-Frenkel *ZF* set theory can be interpreted, e.g. sets of decidable data strings in the alternative Vopenka theory of sets [28], such as: extended sets [3] and multi sets [2]. For these sets, formal axiomatic set theories have been built, i.e. precise descriptions of these sets.

For any abstraction there is a relational structure called a **granule system**. In this system defines: the set of the universe elements of this structure, the atomic sets determined by the universe elements (singletons), relationships like membership of elements to sets, sets conclusion, sets equality. Furthermore, any set is a sum of atomic sets and analogously to standard one, all operations on sets are specified. In any granule system, the *ZF* set theory language is interpreted. However, not all axioms (except from the axioms of the granule system) and not all *ZF* theorems must be met in this system. The granule system, at a higher level of abstraction, specifies knowledge that is inaccurate, uncertain or unclear at a lower level of abstraction. For fifty years, for fuzzy sets, no such system has

been defined and no formal theory has been built. In 1981, Pawlak [15] described the approximated set at a higher level of abstraction as an abstract class of the relation of the equal approximation of sets and proposed to build an axiomatic rough set theory. A solution to this problem is proposed in the Bryniarski's papers [10], [11].

Recently, *the information retrieval IR* in the semantic network, especially in the Semantic Web, usually means looking for a reliable source of this information. So far, information retrieval systems and information interpretations have only indicated semantically the nearest, described in the thesaurus, sources of searched information. However, this is not always the case. Often, when searching for information about an object in a language deviating from the thesaurus, uncertain, unclear or inaccurate knowledge is obtained. Nevertheless, this uncertainty may lead to the unequivocal establishment of sources of knowledge about this object, i.e. precise knowledge. In this way, compliance with the description of the object model is obtained (compatibility with the thesaurus). The situation described above is called **the information disambiguation paradox** of information retrieval [5].

Searching information in the Semantic Web is to find data copies which are:

- one-argument values of attributes – data representing knowledge about some features or types of objects,
- two-argument values of attributes – data representing knowledge about some properties of objects or relations between two objects..

In first case, data are called *concepts*, and in the second one they are called *roles*. To describe concepts and roles, *the Description Logic (DL)* language [1], [4] is used. The *DL* language describing concepts and roles can be extended to some formulas of the first order logic. In the extended language, a *thesaurus* is created, which describes model concepts and roles, while *the ontology* is a language which describes searched concepts and roles. For the searched data described in the ontology and the recommendations (criteria and knowledge) of experts, there may be a certain degree of compliance of these data with the data described in the thesaurus. This is the assessment of the compatibility of data with the thesaurus accepted by experts. The conceiving rule determining the paradox of accuracy appearing here is called **the residuum rule** [5].

This paper presents a perception model of descriptions representing information in semantic networks. In this model, accepted methods to the description perceptions are used, in order to use the residuum rule. It is the perception of references information resources about the object to the degree of compliance of this information with the precisely determined knowledge represented in the set of object descriptions called the thesaurus. Such perceptual system for descriptions will be called **the residuum system**.

The model of information granule systems represented in semantic networks was formulated at the syntactic and semantic level in the papers [7], [9]. Continuing this research, only

the method of inducing information granules by the residuum system will be presented in this paper.

In this paper firstly is presented the semantic network and the perception in the residuum systems in this network. Further is definition of the information granule database and its extension to the information granule system. The perceptual system defined by the information granule database is novel in this paper. At the end there are two examples of such system – rough and fuzzy one.

## II. THE SEMANTIC NETWORK

The semantic network, or the Semantic Web, most commonly is considered to be a graph schema of knowledge representation. It can be identified with an ordered, indexed graph. In the semantic network the vertices and edges are described by some attributes: one or two-argument. In this paper, a more general graphical scheme of knowledge representation is given in which edges can have more than two vertices [7], [8], [9].

*Definition 2.1:* The semantic network is a system:

$$\text{SN} = \langle U, AS, DS \rangle, \quad (1)$$

where:

- $U$  – is a finite set of individual names, object names of represented knowledge (in the Semantic Web it is a set of names which have the Web address). Elements of  $U$  are called **vertices** of the semantic network.
- $DS$  is a family of nonempty sets of vertices descriptions, and also certain systems of these vertices, called **edges**. The number  $n$  is the largest number of vertices in edges.

Let  $\text{card}(U) = n$  and  $U_{gen} = U \cup U^2 \cup \dots \cup U^n$ . Then:

$$AS \subseteq DS \times U_{gen}, \quad (2)$$

Elements of the set  $AS$  are called **assertions**.  $AS$  includes all vertices  $U$ :

$$U = \{x : \text{exists } (ds_k, (x_1, \dots, x_i, \dots, x_k)) \in AS, x = x_i\} \quad (3)$$

When  $ds \in DS$ , then exists a set  $X(ds) \subseteq U_{gen}$ , such that:

$$\{ds\} \times X(ds) = (\{ds\} \times U_{gen}) \cap AS. \quad (4)$$

Such set and its any subset  $X$  is called a *subject  $X$  with description  $ds$*  (shortly: **subject**), and pair  $\langle ds, X \rangle$  is called a *conceiving subject of description  $ds$* . The subject  $X$  with description  $ds$  will be identified with the conceiving subject  $\langle ds, X \rangle$  of description  $ds$ .

For example, let the conceiving subject of description  $ds$  be  $\langle ds, X \rangle$ , where  $R$  is  $k$ -th argument relation such that:

$$\{ds\} \times R = (\{ds\} \times U^k) \cap X(ds). \quad (5)$$

The set  $X(ds)$  is a relation or sum of relations  $R$  defined as above for any number of arguments. About subject  $X(ds)$ , it is said that, it is a *maximum subject* of the description  $ds$ , and about the description  $ds$ , that it is an *instance of the subject*  $X \subseteq X(ds)$ .

$\text{SN}$  is called *full*, if sum of all such sets is equal  $U_{gen}$ .

Any element of  $U_{gen}$  is called *an instance* of the **SN**, and when this element belongs to some subject  $X$  with some description, it is called *an instance occurrence in the SN network*.

Sets  $\{ds\} \times X(ds) = (\{ds\} \times U_{gen}) \cap AS$  are called *attributes* of subjects  $X \subseteq X(ds)$  with descriptions  $ds$ , and these descriptions will be identified with the instance of this attribute. The family of all such subjects  $X(ds)$  is denoted by  $C_0$ . This set is sometimes a division of  $U_{gen}$  set. The elements of  $X(ds)$  are called *an instance occurrence of attribute ds*, and elements of  $AS$  are *assertions*. The following notary agreement is accepted:

For attribute  $ds$  the instance occurrence set about this attribute is denoted as:

- $\|ds\| =_{df} X(ds)$ .
- the assertion occurrence  $(ds, u) \in AS$ , is denoted as  $u : ds$ , eg. instead of '6 < 9' it will be written '(6, 9) : <'.

One-argument relations will be called *concepts*, and at least two-argument relations will be called *roles*, i.e. concepts and roles are subjects of conceiving certain descriptions.

Due to the fact that for any  $X$  subject, uniquely designated by the attribute  $\{ds\} \times X$ , the description  $ds$  corresponds only to one relationship  $R \subseteq X$ ,  $\{ds\} \times R = (\{ds\} \times U^k) \cap X$  with a given number  $k$  of arguments. Therefore, in the further part of this paper, with a fixed number of arguments of these relations, *concepts and roles will be identified with the corresponding descriptions*.

Occurrences of instances with some attribute, occurrence of attributes, subjects with this attribute, concepts, roles and assertions are described in attributive language  $AL$ . Basic syntax and semantic of  $AL$  language are formulated in the paper [1], and the generalized construction of this language is presented in papers [4], [5], [6], [7], [8], [9].

For example, a role *sonhood* connecting a person named *John* with a person named *Simon*, who is his father, leads to assertion:  $\langle sonhood, John, Simon \rangle$ , what can be denoted as:  $sonhood(John, Simon)$  or  $(John, Simon) : sonhood$ .

To join the concept *sonhood* with the time *current year*, we need two assertions  $\langle sonhood, John, Simon \rangle$  and  $\langle sonhood, current\_year \rangle$ , what can be written as a set of descriptions  $\{(John, Simon) : sonhood, (current\_year) : sonhood\}$ .

An assertion which is expressed in a sentence *Eva sits between John and Simon* can be denoted as:  $sit\_between(Eva, John, Simon)$  or  $(Eva, John, Simon) : sit\_between$ . Roles which are functions, in terms of the last component, are called **operations**, for example in the assertion *drive\_to (John, New York)*.

It is significant to notice that in a triple  $(Eva, John, Simon)$  the cyclic inverse of names can be used, and then the following triple is created:  $(John, Simon, Eva)$ , which is also an instance of some role. This new assertion can be expressed in a sentence *John and Simon sit next to Eva* and can be denoted as:  $(John, Simon, Eva) : sit\_nextto$ . The role *sit\_nextto* is **cyclically inverse** to the role *sit\_between*.

When a triple  $(Eva, John, Simon)$ , which is an occurrence of an assertion *sit\_between*, is reduced by

the first component, then a pair  $(John, Simon)$  is also an instance of some assertion, for example expressed in a sentence: *someone sits between John and Simon* –  $(John, Simon) : someone\_sits\_between$ . This role is called **a reduction** of a role *sit\_between*.

Distinguished by experts subsystem of **SN** is denoted as  $SN^+$  in which concepts and roles are considered to be accurate – experts have confidence in this knowledge.  $SN^+ = \langle U^+, AS^+, DS^+ \rangle$  is called a **confidence range** for the **SN**. In the confidence range is  $U_{gen}^+ = U^+ \cup (U^+)^2 \cup \dots \cup (U^+)^n$ . The set  $SN_{tez} = DS^+ \cup SN_{inst}^+$  of all attribute descriptions and instances of these attributes in the  $SN^+$  is called a **thesaurus** of the semantic network **SN** [8], [9].

### III. DESCRIPTION PERCEPTION IN THE RESIDUUM SYSTEMS

In this paper, some aspects of the perceptual proximity theory are used in the context of the proximity of knowledge searched in the semantic Web to the adequate knowledge represented in the thesaurus. A certain view of nearness perception is accepted, combining the basic understanding of perception in psychophysics with the view of the perception described in the Merleau-Ponty paper [13]. This means that the perception of nearness of knowledge about reality to adequate knowledge – and as a result to human knowledge about objects – depends on the signals of sensors, i.e. signals of the senses or measuring systems [12].

But it is known that these signals from measuring systems are received by our senses, and then, as descriptions of objects, are analyzed in the mind. In this approach, our senses are compared to the sampling function. They mimic the impressions describing the features on the numerical values recognized by the mind. Human sensors (senses) collect data samples and measure the physical characteristics of objects in our environment. The physical properties of the object that are read are described and identified with the features of the object. It is our mind that identifies the relations between the values of the features of the object, creating the perception of the detected objects [13]. Object perception is a measure of the adequacy of the information resource that defines this object. As it was written earlier, such a measurement of the adequacy of the information resource will be hereinafter referred to as *a granule of information*. In this sense, by searching for information about certain objects in the Semantic Web, *the perception of object descriptions* and descriptions representing information about these objects are made. In this way a certain set of information granules is obtained [8], [9].

Thus, the definition of algorithms for granular calculations should begin with the definition that is a part of the definition of *the perceptual system*.

#### A. The residuum system

*Definition 3.1:* The system

$$\mathbf{S_P} = \langle S_P, \bullet_P, \rightarrow_P, 0_P, 1_P \rangle \quad (6)$$

is a **residuum system**, where a set  $S_P$  is called *the set of perception values*, and it includes two different elements  $0_P, 1_P$ , called values of *truth* and *false*. An operation  $\bullet_P : S_P \times S_P \rightarrow S_P$  is called *the operation of perception combination*, an operation  $\rightarrow_P : S_P \times S_P \rightarrow S_P$  is called the operation of *perception residuum*.

Operations of the residuum system satisfy following conditions (for any  $z \in S_P$ ):

$$\text{if } z \neq 0_P, \text{ then } (0_P \rightarrow_P z) = 1_P, (z \rightarrow_P 0_P) = 0_P. \quad (7)$$

In addition, there is such an operation  $\sum_P : \wp(S_P) \rightarrow S_P, \wp(S_P) = \{X : X \subseteq S_P\}$ , called *the generalized combination of perception*, such that for any  $x \in S_P$  and for any nonempty disjoint sets  $A, B \subseteq S_P$ :

$$\sum_P \emptyset = 0_P, \quad (8)$$

$$\sum_P \{x\} = x, \quad (9)$$

$$\sum_P (A \cup B) = \sum_P (A) \bullet_P \sum_P (B), \quad (10)$$

Hence, for any  $x, y \in S_P$ :

$$\sum_P \{x, y\} = (x \bullet_P y). \quad (11)$$

A differentiated operation  $(\cdot^d) : S_P \rightarrow S_P$ , such that  $(0_P)^d = 1_P$  and  $(1_P)^d = 0_P$ ,  $(x^d)^d = x$ , is called *the dual value operation* in the system  $\mathbf{S}_P$ , such that if  $x, y \in S_P$  and  $x < y$ , then  $y^d < x^d$ .

- If  $(z \rightarrow_P z) = 1_P$ , then the residuum operation is called *the t-residuum*.
- If  $(z \rightarrow_P z) = 0_P$ , then the residuum operation is called *the s-residuum*.

With the above definitions results:

*Fact 3.1:* Perception combination is a commutative and associative operation.

Let in the residuum system  $\mathbf{S}_P = \langle S_P, \bullet_P, \rightarrow_P, 0_P, 1_P \rangle$ , for the operation  $(\cdot^d)$  of dual value in the system  $\mathbf{S}_P$ , exist such operation  $\sum_{P'} : \wp(S_P) \rightarrow S_P$ , that (for any  $x \in S_P$  and  $A, B \subseteq S_P$ ):

$$\sum_{P'} \emptyset = 0_P, \quad (12)$$

$$\sum_{P'} \{x\} = x, \quad (13)$$

$$\sum_{P'} (A \cup B) = (\sum_{P'} (A)^d \bullet_P \sum_{P'} (B)^d)^d, \quad (14)$$

Then, for any numbers  $x, y \in S_P$  it is assumed that:

$$x \bullet_{P'} y =_{df} (x^d \bullet_P y^d)^d, \quad (15)$$

for any nonempty, disjoint sets  $A, B \subseteq S_P$ :

$$\sum_{P'} (A \cup B) = \sum_{P'} (A) \bullet_{P'} \sum_{P'} (B). \quad (16)$$

Let  $x \rightarrow_{P'} y =_{df} (x^d \rightarrow_P y^d)^d$ .

If  $z \neq 0_{P'}$ , then  $(0_P \rightarrow_{P'} z) = 1_P, (z \rightarrow_{P'} 0_P) = 0_P$ . Therefore:

*Fact 3.2:* The algebra system  $\mathbf{S}_{P'} = \langle S_P, \bullet_{P'}, \rightarrow_{P'}, 0_P, 1_P \rangle$  is the residuum system.

*Definition 3.2:* The residuum system

$\mathbf{S}_{P'} = \langle S_P, \bullet_{P'}, \rightarrow_{P'}, 0_P, 1_P \rangle$  is called **the dual system** for the  $\mathbf{S}_P$  system.

*Fact 3.3:* The operation  $\rightarrow_{P'}$  in the residuum system  $\mathbf{S}_P$  is the s-residuum operation.

**B. T-norm and s-norm systems in the partially ordered set**

*Theorem 3.1:* The algebra system  $\mathbf{S}_t = \langle L, \bullet_t, \rightarrow_t, 0_L, 1_L \rangle$ , is called **the t-norm system** in the set  $L$  partially ordered by the relation  $\leq$ , in which any subset has infimum and supremum, where  $0_L = \inf L$ , and  $1_L = \sup L$ . It is the residuum system, if the operation  $\bullet_t : L \times L \rightarrow L$ , called *the t-norm* in the  $L$ , satisfies following conditions (for any numbers  $w, x, y, z \in L$ ):

- boundary conditions

$$0_L \bullet_t y = 0_L, y \bullet_t 1_L = y \quad (17)$$

- uniform value increase, monotonicity

$$x \bullet_t y \leq z \bullet_t y, \text{ when } x \leq z \quad (18)$$

- uniform value limitation

$$w \leq x \bullet_t y \leq z, \text{ when } w \leq x \leq z \text{ or } w \leq y \leq z \quad (19)$$

- commutativity

$$x \bullet_t y = y \bullet_t x \quad (20)$$

- associativity

$$x \bullet_t (y \bullet_t z) = (x \bullet_t y) \bullet_t z \quad (21)$$

- and

$$\text{exist } x \rightarrow_t y = \sup\{t \in L : x \bullet_t t \leq y\}. \quad (22)$$

Such described operation  $\rightarrow_t : L \times L \rightarrow L$  is *the t-residuum* in the set  $L$ .

Let operation  $\cdot^d : L \rightarrow L$  be an operation of the dual values in the  $\mathbf{S}_t$  system, then the system  $\mathbf{S}_s = \langle L, \bullet_s, \rightarrow_s, 0_L, 1_L \rangle$  is defined as follows (for any numbers  $x, y \in L$ ):

$$x \bullet_s y = (x^d \bullet_t y^d)^d, \quad (23)$$

$$x \rightarrow_s y = (x^d \rightarrow_t y^d)^d, \quad (24)$$

Then:

*Theorem 3.2:* In the system  $\mathbf{S}_s = \langle L, \bullet_s, \rightarrow_s, 0_L, 1_L \rangle$ , the following conditions are satisfied (for any numbers  $w, x, y, z \in L$ ):

- boundary conditions

$$0_L \bullet_s y = y, y \bullet_s 1_L = 1_L \quad (25)$$

- uniform value increase, monotonicity

$$x \bullet_s y \leq z \bullet_s y, \text{ when } x \leq z \quad (26)$$

- uniform value limitation

$$w \leq x \bullet_s y \leq z, \text{ when } w \leq x \leq z \text{ or } w \leq y \leq z \quad (27)$$

- commutativity

$$x \bullet_s y = y \bullet_s x \quad (28)$$

- associativity

$$x \bullet_s (y \bullet_s z) = (x \bullet_s y) \bullet_s z \quad (29)$$

- and

$$x \rightarrow_s y = \inf\{t \in L : y \leq x \bullet_s t\}. \quad (30)$$

*Definition 3.3:* The system  $\mathbf{S}_s = \langle L, \bullet_s, \rightarrow_s, 0_L, 1_L \rangle$  is called **the s-norm system** in the set  $L$  partially ordered by the relation  $\leq$ , and the operation  $\bullet_s : L \times L \rightarrow L$  is called *the s-norm* in the  $L$  for the operation  $(\cdot^d)$  of dual values in the system  $\mathbf{S}_t$ .

*Example 3.1:* Let range of numbers  $L = [0, 1]$  be ordered by the relation  $\leq$ . The operation  $\bullet_t : [0, 1] \times [0, 1] \rightarrow [0, 1]$ , for any  $x, y \in [0, 1]$ , is the t-norm and is defined by formula:

$$x \bullet_t y = \inf\{x, y\} = \min\{x, y\}. \quad (31)$$

Its generalized form determines the formula, for any set  $X \subseteq [0, 1]$

$$\sum_t X = \inf X. \quad (32)$$

It can be determined that:

$$x \rightarrow_t y = \sup\{t \in [0, 1] : \min\{x, t\} \leq y\}. \quad (33)$$

*Example 3.2:* Let range of numbers  $L = [0, 1]$  be ordered by the relation  $\leq$ . The operation  $\bullet_s : [0, 1] \times [0, 1] \rightarrow [0, 1]$  for any  $x, y \in [0, 1]$ , is the s-norm and is defined by formula:

$$x \bullet_s y = \sup\{x, y\} = \max\{x, y\} \quad (34)$$

Its generalized form determines the formula, for any set  $X \subseteq [0, 1]$ :

$$\sum_s X = \sup X. \quad (35)$$

It can be determined that:

$$\begin{aligned} x \rightarrow_s y &= 1 - (1 - x) \rightarrow_t (1 - y) = \\ &= 1 - \sup\{t \in [0, 1] : \min\{1 - x, t\} \leq 1 - y\} = \\ &= \inf\{t \in [0, 1] : y \leq x \bullet_s t\}. \end{aligned} \quad (36)$$

Having t-norm and s-norm systems can be determined:

*Definition 3.4:* The system  $\mathbf{S}_{\text{logic}} = \langle L, \bullet_s, \rightarrow_t, 0_L, 1_L \rangle$  is called **the logical residuum system**.

#### IV. THE INFORMATION GRANULE DATABASE

Firstly, the definition of the perceptual system is given in order to define the information granule database. Then, it is shown how this database is induced by this system.

##### A. The perceptual system

*Definition 4.1:* For the semantic network  $\mathbf{SN} = \langle U, AS, DS \rangle$ , a **perceptual object** is an instance which was given a certain value in the residuum system  $\mathbf{S}_{\text{logic}} = \langle L, \bullet_s, \rightarrow_t, 0_L, 1_L \rangle$ , i.e. perceptual objects are elements of some set  $O = U_{gen}$ , where the set  $L_0 \subseteq L$  is a set of values of instances in the residuum system  $\mathbf{S}_{\text{logic}}$ .

Giving certain values in the  $\mathbf{S}_{\text{logic}}$  residuum system also has some interpretation in  $\mathbf{S}_{\text{logic}}$ . This interpretation is defined by the definition:

*Definition 4.2:* A **probe function** or a **perception** is a function that  $\phi : O \rightarrow L_0$  represents a feature of a perceptual object [23], [26].

Further, for the  $\mathbf{SN}$ , it is assumed that any description  $\phi \in DS$  corresponds to a certain perception  $\phi : O \rightarrow L_0$  determined by this description  $\phi$ .

Extending the perceptual system definition [26], for the concept of the logical residuum system  $\mathbf{S}_{\text{logic}}$ , it is assumed that:

*Definition 4.3:* A **perceptual system**  $\mathbf{PS} = \langle O, F, \mathbf{S}_{\text{logic}} \rangle$  consists of a nonempty set  $O$  of sample perceptual objects and the set  $F$  of chosen perceptions  $\phi : O \rightarrow L_0$ , called *the perceptions of the PS system*. Elements of the  $L_0$  are called then *the perception degrees*.

##### B. The granule information database induced by the perceptual system

*Definition 4.4:* The system:

$$\mathbf{G}_{\text{base}} = \langle G, \{ \}_G, \cup_G, \subseteq_G, =_G, 0_G, 1_G, G_{inst}, G_{set}, G_0 \rangle, \quad (37)$$

is called **the granule information database**, where elements of the set  $G$  are called *granules*,  $G_{inst}$  is a set of *instance granules*,  $G_{set}$  is a set of *set granules* and  $G_0$  is a set of *singletons of granules*, a granule  $0_G$  is called an *empty granule* and a granule  $1_G$  is called a *full granule*, the operations  $\{ \}_G, \cup_G$  and the relations  $\subseteq_G, =_G$  are defined by following conditions:

$$G_0 \subseteq G_{set}, \quad (38)$$

$$G = G_{inst} \cup G_{set}, \quad (39)$$

$$G_{inst} \cap G_{set} = \emptyset. \quad (40)$$

There is some function  $\{ \}_G : G_{inst} \rightarrow G_0$ , such that

$$\text{for any } x \in G_{inst}, \{x\}_G \subseteq_G 1_G, \text{ additionally, if } \{x_0\}_G \subseteq_G \{x\}_G, \text{ then } \{x_0\}_G =_G \{x\}_G, \quad (41)$$

$$G_0 = \{x \in G_{set} : \exists z \in G_{inst} (x = \{z\}_G)\}. \quad (42)$$

There is some function  $\cup_G : \wp(G_{set}) \rightarrow G_{set}$ , such that

$$\cup_G \emptyset = 0_G, \quad (43)$$

$$\cup_G G_{set} = 1_G, \quad (44)$$

$$\cup_G \{z\} = z, \text{ for } z \in G_{set}, \quad (45)$$

$$x \subseteq_G x, \quad (46)$$

$$0_G \subseteq_G x, \quad (47)$$

$$\text{If } x \neq 0_G, \text{ then, it is not true that } x \subseteq_G 0_G, \quad (48)$$

$$y = \cup_G \{x \in G_0 : x \subseteq_G y\}, \quad (49)$$

for any  $x, y \in G_{set}, (x =_G y) \Leftrightarrow_{df}$

$$\cup_G \{z \in G_0 : (z \subseteq_G x) \Leftrightarrow (z \subseteq_G y)\} = 1_G. \quad (50)$$

It is assumed, in the sense of the set theory, that the  $Inst : U_{gen} \rightarrow G_{inst}$  function assigns a certain instance

granule to each instance. Any such function allows to enter a notation agreement for the granule  $Inst(\langle v_1, v_2, \dots, v_k \rangle)$ , when  $\langle v_1, v_2, \dots, v_k \rangle \in U^k$ :

$$\begin{aligned} (Inst)\langle x_1, x_2, \dots, x_k \rangle &= Inst(\langle v_1, v_2, \dots, v_k \rangle) \\ \text{iff } \langle v_1, v_2, \dots, v_k \rangle &\in U^k \text{ and} \\ x_i &= Inst(v_i), \text{ for } i = 2, \dots, k. \end{aligned} \quad (51)$$

**Theorem 4.1:** Let, for any perceptual system  $\mathbf{PS} = \langle O, F, \mathbf{S}_{\text{logic}} \rangle$ ,  $\mathbf{S}_{\text{logic}} = \langle L, \bullet_s, \rightarrow_t, 0_L, 1_L \rangle$ , symbols  $G, \{\}_G, \cup_G, \subseteq_G, =_G, 0_G, 1_G, G_{inst}, G_{set}, G_0$  be interpreted as follows (for any  $\phi_1, \phi_2 \in G_{set}, \phi \in G_0$ ,  $C_0$  is a family of maximum subjects in the semantic network  $\mathbf{SN}$ ):

$$G_{inst} = U_{gen} \times L_0 \cup C_0, \quad (52)$$

where  $C_0 \subseteq \wp(U_{gen})$  and  $L_0 = \{r : \text{exists } a \in U_{gen} \text{ and exists } \phi \in F \text{ such that } r = \phi(a)\}$ ,

$$G_{set} = F, \quad (53)$$

$$G = G_{inst} \cup G_{set}, \quad (54)$$

$$\phi_1 \subseteq_G \phi_2 \text{ iff for any } x \in L, \phi_1(x) \rightarrow_t \phi_2(x) = 1_L, \quad (55)$$

$$\{(a, r)\}_G = \mu \in F, \quad (56)$$

where  $\mu$  is such perception that for  $(a, r) \in G_{inst}, \mu(a) = r \neq 0_L$ , and  $\mu \subseteq_G 1_G$ , additionally, if  $\mu_0 \in F$  and  $\mu_0 \subseteq_G \mu$ , then  $\mu_0 = \mu$ ,

$$\{K\}_G = \mu \in F, \quad (57)$$

where  $\mu$  is such perception that for  $K \in C_0, \mu(a) = 1_L$  for any  $a \in K, \mu(a) \neq 0_L$  for any  $a \notin K$  and  $\mu \subseteq_G 1_G$ , additionally, if  $\mu_0 \in F$  and  $\mu_0 \subseteq_G \mu$ , then  $\mu_0 = \mu$ ,

$$\begin{aligned} G_0 &= \{\phi \in F : \text{exists } u \in G_{inst}, \\ &\text{such that } \phi = \{u\}_G\}. \end{aligned} \quad (58)$$

Additionally, if  $A \subseteq F$ , then:

$$(\cup_G A)(x) = \sum_P \{y \in L : y = \phi(x) \wedge \phi \in A\}, \quad (59)$$

$$\begin{aligned} (\phi_1 =_G \phi_2) &\Leftrightarrow_{df} \cup_G \{\phi \in G_0 : (\phi \subseteq_G \phi_1) \Leftrightarrow \\ &\Leftrightarrow (\phi \subseteq_G \phi_2)\} = 1_G. \end{aligned} \quad (60)$$

Then, the system  $\langle G, \{\}_G, \cup_G, \subseteq_G, =_G, 0_G, 1_G, G_{inst}, G_{set}, G_0 \rangle$  is the granule information database.

**Definition 4.5:** The granule database described in above theorem is called the granule information database induced by the perceptual system  $\mathbf{PS}$ .

## V. EXTENDING THE INFORMATION GRANULE DATABASE TO THE INFORMATION GRANULE SYSTEM

**Definition 5.1:** For any  $x \in G_{inst}$  and  $y \in G_{set}$ ,

$$x \in_G y \text{ iff } \{x\}_G \subseteq_G y. \quad (61)$$

The relation  $\in_G$  is called the **relation of belonging instance granule to the granule set**.

Hence, and from the conditions describing the information granule database:

**Theorem 5.1:**

$$\{x \in G_{inst} : x \in_G 0_G\} = \emptyset, \quad (62)$$

$$\{x \in G_{inst} : \exists z \in G_0 (x \in_G z)\} = G_{inst}, \quad (63)$$

$$\{z \in G_{set} : \exists x \in G_{inst} (x \in_G z)\} = G_{set}, \quad (64)$$

$$y = \cup_G \{\{x\}_G : x \in_G y\}, \quad (65)$$

**Theorem 5.2:** Granules  $y^G, y \cap_G z, y \cup_G z$ , and  $y \setminus_G z$  exist in the granule database, and are defined by formulas:

$$y^G = \cup_G \{z \in G_0 : \exists x \in G_{inst} (x \in_G z \wedge \neg x \in_G y)\}, \quad (66)$$

$$\begin{aligned} y \cap_G z &= \cap_G \{y, z\} = \cup_G \{t \in G_0 : \\ &\exists x \in G_{inst} (x \in_G t \wedge x \in_G y \wedge x \in_G z)\}, \end{aligned} \quad (67)$$

$$\begin{aligned} y \cup_G z &= \cup_G \{y, z\} = \cup_G \{t \in G_0 : \\ &\exists x \in G_{inst} (x \in_G t \wedge (x \in_G y \vee x \in_G z))\}, \end{aligned} \quad (68)$$

$$\begin{aligned} y \setminus_G z &= \cup_G \{t \in G_0 : \\ &\exists x \in G_{inst} (x \in_G t \wedge x \in_G y \wedge \neg x \in_G z)\}. \end{aligned} \quad (69)$$

In order to unify the expressions of the attributive language  $AL$  in various theories of information granules, formulated in theories: set theory, probability theory, possible data sets in the evidence systems, fuzzy set theory, rough sets theory and shadowed sets theory, the expressions of the attributive language are assumed to be interpreted in a chosen relational system  $\mathbf{G}$  [8], [9], [11] given below. A distinction is made between the set of granule instances and the set of granule set instances. In the first set, attribute instances are interpreted and in the second - set of instances (concepts and roles). The important thing is that the granule set instances determine sets of instances. Granule instances are interpreted as elements of granule set instances, analogically to some classical  $\mathbf{G}^+$  algebra.

**Definition 5.2:** Let the granule system for the attributive language be:

$$\begin{aligned} \mathbf{G} &= \langle G, M_G, \cup_G, \cap_G, \setminus_G, 'G, \in_G, \subseteq_G, \\ &=_{=G}, 0_G, 1_G, G_{inst}, G_{set}, G_0 \rangle. \end{aligned} \quad (70)$$

where:

- $G = G_{inst} \cup G_{set}$  is a sum of sets:  $G_{inst}$  – is a set of granule instances and  $G_{set}$  is a set of granule set instances,
- $M_G$  is a set of functions  $m_G : G \rightarrow G$ ,
- operations  $\cup_G, \cap_G$  are generalized operations of sum and product described on the subsets of the granules family  $G$ ,
- $\setminus_G$  is an operation of granules difference,
- $'G$  is an operation of granules closure,
- for an empty set, a value of these generalized operations is an empty granule  $0_G$  and for the  $G$  set it is a full granule  $1_G$ ,

- $\in_G$  is a relation of being a granule element,
- $\subseteq_G$  is a relation of granules inclusion for instance set granules,
- $=_G$  is a relation of granules closeness for instance set granules,
- $G_0$  is a set of chosen granules.

$G_0$  is a set of granules called *granules of instances singletons* such that there is some function  $\{\}_G : G_{inst} \rightarrow G_0$  and  $G_0 = \{x \in G : \exists z \in G_{inst}(x = \{z\}_G)\}$ .

## VI. EXAMPLES OF THE GRANULE SYSTEM

Let consider two examples of the granule system the rough and the fuzzy granule databases.

### A. The rough granule database

Let in the system  $\langle U_{gen}, C \rangle$ , where  $C$  is a partition of  $U_{gen}$ , operations be defined (for any  $X \subseteq U_{gen}$ ):

$$C^-(X) = \cup\{K \in C : K \subseteq X\}, \quad (71)$$

$$C^+(X) = \cup\{K \in C : K \cap X \neq \emptyset\}, \quad (72)$$

Any sets  $X, Y \subseteq U_{gen}$  are indiscernibility, what is written:  $X \sim Y$  iff

$$C^-(X) = C^-(Y), \quad (73)$$

$$C^+(X) = C^+(Y). \quad (74)$$

The relation  $\sim$  is a relation of equivalence. The abstraction classes  $[X]_{\sim}$  of this relation for the representative  $X$  is denoted as  $X_C$ .  $\emptyset_C$  is denoted by  $0_C$  and  $(U_{gen})_C$  is denoted by  $1_C$ .

The abstract classes of the relations  $\sim$  are called **rough sets** in the system  $\langle U_{gen}, C \rangle$ .

The set inclusion relations and the rough membership relation [10] is defined as follows (for any  $X, Y \subseteq U_{gen}$ ):

$$X \subseteq_C Y_C \text{ iff } C^-(X) \subseteq C^-(Y) \text{ and } C^+(X) \subseteq C^+(Y), \quad (75)$$

$$X \in_C Y_C \text{ iff } X \neq \emptyset \text{ and exists such } K \in C, \text{ that } X \subseteq K, C^-(X) \subseteq C^-(Y) \text{ and } K \subseteq C^+(Y). \quad (76)$$

The expression  $X \in_C Y_C$  is read:  $X$  is an element of the rough set  $Y_C$ . Hence:

$$X \in_C Y_C \text{ iff } X \neq \emptyset \text{ and exists such } K \in C, \text{ that } X \subseteq K \text{ and } X \subseteq_C Y_C. \quad (77)$$

Intuitively, due to the fact that the description of  $x \in Y$  cannot be precisely determined, this description is interpreted as follows: indistinguishable from  $x$  elements of the indistinguishable elements from the sets  $Y$ . With the relation  $\in_C$ , the conclusion of rough sets can also be defined. For any  $Y \subseteq U_{gen}$ ,

$$X_C \subseteq_C Y_C \text{ iff for any } Z \subseteq U_{gen}, \text{ if } Z \in_C X_C, \text{ then } Z \in_C Y_C. \quad (78)$$

Using the theorems given by Bryniarski [10], [11], in the family approximate sets, analogically to the classical set theory, the following operations can be defined: the addition  $\cup_C$ , the multiplication  $\cap_C$ , the difference  $\setminus_C$  and the complement  $'_C$  of the rough sets.

For any rough sets  $X_C, Y_C$ , and any  $Z \subseteq U_{gen}$ ,

$$Z \in_C X_C \cup_C Y_C \text{ iff } Z \in_C X_C \text{ or } Z \in_C Y_C, \quad (79)$$

$$Z \in_C X_C \cap_C Y_C \text{ iff } Z \in_C X_C \text{ and } Z \in_C Y_C, \quad (80)$$

$$Z \in_C X_C \setminus_C Y_C \text{ iff } Z \in_C X_C \text{ and not } Z \in_C Y_C, \quad (81)$$

$$Z \in_C (X_C)'_C \text{ iff } Z \in_C (U_{gen})_C \setminus_C X_C, \quad (82)$$

Operations  $\cup_C, \cap_C$  can be generalized and used in the same way as in the set theory.

If the system:

$$\mathbf{G}_{rough} = \langle G, M_G, \cup_G, \cap_G, \setminus_G, '^G, \{\}_G, \in_G, \subseteq_G, =_G, 0_G, 1_G, G_{inst}, G_{set}, G_0 \rangle, \quad (83)$$

is interpreted as follows:

$$G_{set} =_{df} \{X_C : X \subseteq U_{gen}\}, \quad (84)$$

$$G_{inst} =_{df} \{X \subseteq U_{gen} : X = \{x\} \text{ or } X \in C\}, \quad (85)$$

$$G =_{df} G_{inst} \cup G_{set}, \quad (86)$$

$$\cup_G =_{df} \cup_C, \quad (87)$$

$$\cap_G =_{df} \cap_C, \quad (88)$$

$$\setminus_G =_{df} \setminus_C, \quad (89)$$

$$'^G =_{df} ' _C, \quad (90)$$

$$\in_G =_{df} \in_C, \quad (91)$$

$$\subseteq_G =_{df} \subseteq_C, \quad (92)$$

$$=_G =_{df} =, \quad (93)$$

$$0_G =_{df} \emptyset_C = \{\emptyset\}, \quad (94)$$

$$1_G =_{df} (U_{gen})_C, \quad (95)$$

An operation  $\{\}_G : G_{inst} \rightarrow G_0$  such that for any  $X \in G_{inst}$ ,  $\{X\}_G = X_C$ ,  $G_0 = \{X_C : X \in G_{inst}\}$ , and the set of operations  $M_G$  is empty.

Then the system  $\mathbf{G}_{base}$  (equation 37 from definition 4.4) is a granule information database, and the elements of  $G$  are called **the approximate granules**. Moreover, the system  $\mathbf{G}_{rough}$  is **the rough granule system**.

### B. The fuzzy granule database

Let  $\mathbf{PS} = \langle O, F, \mathbf{S}_{logic} \rangle$  be a perceptual system, in which  $O = U_{gen}$ , and the instance values  $L_0 \subseteq [0, 1]$  are in the residuum system  $\mathbf{S}_{logic} = \langle [0, 1], \bullet_s, \rightarrow_t, \{0, 1\} \rangle$ . In the range  $[0, 1]$ , the s-norm  $\bullet_s : [0, 1] \times [0, 1] \rightarrow [0, 1]$  is defined by formula (for any  $x, y \in [0, 1]$ ):

$$x \bullet_s y = \sup\{x, y\} = \max\{x, y\}. \quad (96)$$

Its generalized form determines the formula (for any set  $X \subseteq [0, 1]$ ):

$$\sum_s X = \sup X. \quad (97)$$

It can be determined that:

$$x \rightarrow_t y = \sup\{t \in [0, 1] : \min\{x, t\} \leq y\}. \quad (98)$$

$F$  is a set of symbols  $\mu_A$  of a function  $\mu_A : U_{gen} \rightarrow [0, 1]$ , and also a symbol of some fuzzy sets [29], [30], [31], described for any  $A \subseteq U_{gen}$  as follows (for any  $x \in K$ ) [17]:

$$[x] = K \Leftrightarrow K \in C \text{ and } x \in K, \quad (99)$$

$$\mu_A(x) = |A \cap [x]|/|[x]|, \quad (100)$$

i.e.  $F = \{\mu_Y : Y \subseteq U_{gen}\}$ , where  $|X|$  denotes the cardinality of the  $X$ .

Hence the result:

*Theorem 6.1:* There are (for  $A, B \subseteq U_{gen}$ ):

$$C^-(A) = \{x \in U_{gen} : \mu_A(x) = 1\}, \quad (101)$$

$$C^+(A) = \{x \in U_{gen} : \mu_A(x) > 0\}, \quad (102)$$

$$A_C = B_C \Leftrightarrow \text{for any } x \in U_{gen}, \mu_A(x) = \mu_B(x) \Leftrightarrow \mu_A = \mu_B, \quad (103)$$

$$A_C \subseteq_C B_C \Leftrightarrow \text{for any } x \in U_{gen}, \mu_A(x) \rightarrow_t \mu_B(x) = 1, \quad (104)$$

$$\mu_A(x) \rightarrow_t \mu_B(x) = 1 \Leftrightarrow \mu_A(x) \leq \mu_B(x), \quad (105)$$

$$A_C \subseteq_C B_C \Leftrightarrow \text{for any } x \in U_{gen}, \mu_A(x) \leq \mu_B(x). \quad (106)$$

The perceptual system  $\mathbf{PS} = \langle O, F, \mathbf{S}_{logic} \rangle$ , induces the **fuzzy granule database**  $\mathbf{G}_{base}$  (equation 37 from definition 4.4), where symbols  $G, \{ \}_G, \cup_G, \subseteq_G, =_G, 0_G, 1_G, G_{inst}, G_{set}, G_0$  are interpreted as follows:

$$G_{inst} = U_{gen} \times L_0 \cup C, \quad (107)$$

where  $C$  is a partition of  $U_{gen}$  and  $L_0 = \{r : \text{exists } u \in U_{gen} \text{ and exists } \mu \in F \text{ such that } r = \mu(u)\}$ ,

$$G_{set} = F, \quad (108)$$

$$G = G_{inst} \cup G_{set}, \quad (109)$$

$$\mu_1 \subseteq_G \mu_2 \text{ iff for any } x \in U_{gen}, \mu_1(x) \rightarrow_t \mu_2(x) = 1, \quad (110)$$

$$\{(a, r)\}_G = \mu \in F, \quad (111)$$

where  $\mu$  is such perception that  $\mu(a) = r \neq 0$  and  $\mu \subseteq_G 1_G$ , additionally, if  $\mu_0 \in F$  and  $\mu_0 \subseteq_G \mu$ , then  $\mu_0 = \mu$ ,

$$\{K\}_G = \mu \in F, \quad (112)$$

where  $\mu$  is such perception that for  $K \in C, \mu(a) = 1$  for any  $a \in K, \mu(a) \neq 0$  for any  $a \notin K$  and  $\mu \subseteq_G 1_G$ , additionally, if  $\mu_0 \in F$  and  $\mu_0 \subseteq_G \mu$ , then  $\mu_0 = \mu$ ,

$$G_0 = \{\phi \in F : \text{exists } u \in G_{inst}, \text{ such that } \phi = \{u\}_G\}, \quad (113)$$

Additionally, if  $A \subseteq F$ , then

$$(\cup_G A)(x) = \sup\{y \in [0, 1] : y = \mu(x) \wedge \mu \in A\}, \quad (114)$$

$$(\mu_1 =_G \mu_2) \Leftrightarrow_{df} \cup_G \{\mu \in G_0 : (\mu \subseteq_G \mu_1) \Leftrightarrow (\mu \subseteq_G \mu_2)\} = 1_G. \quad (115)$$

## VII. CONCLUSION

Presented application of abstraction methods in creating concepts allows to describe and solve more complex problems of knowledge representation in the semantic network, especially in the Semantic Web. Following issues are presented in this paper:

- A semantic network having a more general graph representation of the knowledge representation has been specified, i.e. one in which the edges of the network can have more than two vertices [5], [7], [8]. Roles in attribute language  $AL$  can join more than two vertices of the network. In addition, all currently used methods of knowledge representation can be implemented in a certain semantic network understood as in this paper.
- The theory of information granule databases in the semantic network has been formulated, in which axioms meet the standard theorems of the set theory defining the concept of a set. The model of this theory is the granule system.
- It has been shown that very complex constructs of the interpretation of the  $AL$  language expressions in the granule systems [9], at a higher level of abstraction, can be simplified by reducing them to interpretation in the granule databases induced by the perceptual system.
- Due to the fact that any granules are sums of singletons, calculations in granule systems can be simplified by performing them only on certain selected representatives of the elements of these granules. That allows to implement such computational procedures for the most frequently occurring in the processing of knowledge large data sets represented in the Semantic Web.

In further work presented information granule system will be designed also for extended sets, multiset, Borel field of sets and the system of conceiving will be defined.

## REFERENCES

- [1] Baader F., Calvanese D., McGuinness D., Nardi D., Patel-Schneider P. (eds.): The Description Logic. Handbook Theory, Implementation and Application. Cambridge University Press, Cambridge, 2003.
- [2] Blizard W. D.: Multiset Theory. Notre Dame Journal of Formal Logic, vol. 30, Number 1, pp. 36-66, 1989.
- [3] Blass, A. C., Childs D. L.: Axioms and Models for an Extended Set Theory. University of Michigan, Mathematics Dept: 2011.
- [4] Bobillo F., Straccia U.: Fuzzy Description Logics with general t-norms and datatypes, Fuzzy Sets Systems, vol. 160(23), pp. 3382-3402, 2009.
- [5] Bryniarska, A.: The Paradox of the Fuzzy Disambiguation in the Information Retrieval. (IJARAI) International Journal of Advanced Research in Artificial Intelligence, pp. 55-58, Volume 2 No 9, 2013.
- [6] Bryniarska A.: The Model of Possible Web Data Retrieval. Proceedings of 2nd IEEE International Conference on Cybernetics CYBCONF 2015, pp. 348-353, 2015.

- [7] Bryniarska A., Bryniarski E.: Rough search of vague knowledge. In: G. Wang, A. Skowron, Y. Yao, D. Slezak, L. Polkowski (eds.), *Thriving Rough Sets-10th Anniversary - Honoring Professor Zdzislaw Pawlak's Life and Legacy & 35 years of Rough Sets*, Studies in Computational Intelligence, Springer, Berlin Heidelberg New York, pp.283-310, 2017.
- [8] Bryniarska A.: Autodiagnosis of Information Retrieval on the Web as a Simulation of Selected Processes of Consciousness in the Human Brain. In: *Biomedical Engineering and Neuroscience*, W. P. Hunek, S. Paszkiel eds., *Advances in Intelligent Systems and Computing* 720, pp. 111-120, Springer, 2018.
- [9] Bryniarska A.: Certain information granule system as a result of sets approximation by fuzzy context, *International Journal of Approximate Reasoning*, Volume 111, pp. 1-20, August 2019, in press.
- [10] Bryniarski E.: A calculus of rough sets of the first order. *Bull. Pol. Ac.: Math.* 37, pp. 109-136, 1989.
- [11] Bryniarski E.: Formal conception of rough sets. *Fund. Infor.* 27(2-3), pp.103-108, 1996.
- [12] Fahlé M., Poggio T.: *Perceptual Learning*. Cambridge, MA: The MIT Press, 2002.
- [13] Merleau-Ponty M.: *Phenomenology of Perception*. Paris and New York: Smith, Callimard, Paris and Routledge & Kegan Paul, 1945.
- [14] Moore R.: *Interval Analysis*, Prentice-Hall, Englewood Clifis, NJ, 1966.
- [15] Pawlak Z.: Rough sets. *Intern. J. Comp. Inform. Sci.* 11, pp. 341-356, 1982.
- [16] Pawlak Z.: *Rough Sets. Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishers, Dordrecht, 1991.
- [17] Pawlak Z, Skowron A.: Rough membership function, in: R. E Yeager, M. Fedrizzi and J. Kacprzyk (eds.), *Advances in the Dempster-Schafer of Evidence*, Wiley, New York, 251-271, 1994.
- [18] Pawlak Z., Skowron A.: Rudiments of rough sets. *Information Sciences*, 177,1, 1, pp. 3-27, 2007.
- [19] Pawlak Z., Skowron. A.: Rough sets and Boolean reasoning. *Information Sciences*, 177, 1, pp. 41-73, 2007.
- [20] Pedrycz W.: Shadowed sets: representing and processing fuzzy sets, *IEEE Transactions on Systems, Man, and Cybernetics - Part B* 28 pp. 103-109, 1998.
- [21] Pedrycz W.: *Knowledge-Based Clustering: From Data to Information Granules*, J. Wiley, Hoboken, NJ, 2005.
- [22] Pedrycz W.: Allocation of information granularity in optimization and decision-making models: towards building the foundations of Granular Computing. *Eur J Oper Res* 232(1),pp. 137-145, 2014. doi:10.1016/j.ejor.2012.03.038
- [23] Peters J. F., Skowron A., Stepaniuk J.: Nearness of objects: Extension of approximation space model. *Fundamenta Informaticae*, vol. 79, no. 3-4, pp. 497-512, 2007.
- [24] Peters J. F.: Discovery of perceptually near information granules. In: *Novel Developments in Granular Computing: Applications of Advanced Human Reasoning and Soft Computation*, J. T. Yao, Ed. Hersey, N.Y., USA: Information Science Reference, 2009.
- [25] Peters J. F., Ramanna S.: Affinities between perceptual granules: Foundations and Perspectives. In: *Human-Centric Information Processing Through Granular Modelling*, A. Bargiela and W. Pedrycz, Eds. Berlin: Springer-Verlag, pp. 49-66, 2009.
- [26] Peters J. F., Wasilewski P.: *Foundations of near sets*. Elsevier Science, vol. 179, no. 1, pp. 3091-3109, 2009.
- [27] Tsichritzis D. C., Lochovsky F.: *Data models*, Published by Prentice Hall, Inc. Englewood Clis, New Jersey, USA, 1982.
- [28] Vopenka P.: *Mathematics in the Alternative Set Theory*. Leipzig: Teubner, 1979.
- [29] Zadeh L.A.: Fuzzy sets. *Inf Control* 8(3): pp. 338-353, 1965.
- [30] Zadeh L.A.: Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, *Fuzzy Sets and Systems* 90, pp. 111-117, 1997.
- [31] Zadeh L.A.: Toward a generalized theory of uncertainty (GTU) an outline, *Information Sciences* 172, pp. 1-40, 2005.



## Improving Real-Time Performance of U-Nets for Machine Vision in Laser Process Control

Przemysław Dolata

Wrocław University of Science and Technology, ul.  
Wyb. Wyspiańskiego 27,  
50-370 Wrocław, Poland  
Email: przemyslaw.dolata@pwr.edu.pl

Jacek Reiner

Wrocław University of Science and Technology  
ul. Wyb. Wyspiańskiego 27,  
50-370 Wrocław, Poland  
Email: jacek.reiner@pwr.edu.pl

**Abstract**—Many industrial machine vision problems, particularly real-time control of manufacturing processes such as laser cladding, require robust and fast image processing. The inherent disturbances in images acquired during these processes makes classical segmentation algorithms uncertain. Among many convolutional neural networks introduced recently to solve such difficult problems, *U-Net* balances simplicity with segmentation accuracy. However, it is too computationally intensive for usage in many real-time processing pipelines.

In this work we present a method of identifying the most informative levels of detail in the *U-Net*. By only processing the image at the selected levels, we reduce the total computation time by 80%, while still preserving adequate quality of segmentation.

### I. INTRODUCTION

SEGMENTATION of complex, noisy images is a core problem in many industrial applications of machine vision, especially in monitoring and control of laser additive manufacturing processes, such as laser cladding [1]. Where classical image processing algorithms cannot provide necessary robustness (against, for example, plasma emissions or powder scattering), machine-learning-based solutions are applied – recently, convolutional neural networks in particular. However, they are notoriously computationally heavy. For off-line applications this issue can be trivially solved with using more compute power, but in some on-line, real-time applications it is a critical problem. If the process state changes rapidly, any delay in its measurement degrades performance of the control algorithm.

*U-Net* is a well-known and proven convolutional neural network architecture for image segmentation [2]. Its distinguishing property is a highly modular, symmetric, dual-path structure. In the “down” path, which comprises blocks of max-pooling and convolution layers, features are being extracted from progressively smaller inputs. Those blocks can

be thought of as observing the input at progressively smaller scales. As a result, they produce feature maps with gradually more contextual information, but less spatial resolution. On the other hand, the “up” path integrates the high-context but low-resolution feature maps with intermediate levels of low-context but high-resolution information. This allows producing highly detailed segmentations for objects of different scales.

Training the *U-Net* on laser cladding monitoring images is a relatively straightforward task, even with a small amount of annotated data. The baseline configuration as described by Ronneberger *et al.* [2] outputs segmentations of satisfactory quality without the need to apply any tricks or problem-specific tuning. However, the time of a single image inference, on our in-house hardware, is approximately 250ms. This is unacceptable for any on-line processing purpose – especially for real-time control.

The simplest yet very effective way to decrease processing time is to reduce the size of the input images. This might have an additional benefit of reducing the cost of data acquisition, or allowing higher processing frame rates. A more advanced method would be to downsample the images in the “down” path earlier, skipping some detail scales during inference if the information they contain does not significantly contribute to the overall segmentation quality. However, it is difficult to determine a priori, at which scale should the input be observed and at which intermediate scales should it be processed. Intuitively, this depends on the specific characteristics of a particular problem. Detecting large objects might require more context – hence, deeper “down” path – than small ones. On the other hand, segmenting objects with fuzzy boundaries might not benefit from very high-resolution features as much as when objects have very clear and detailed edges.

In this study, we present a method of determining which blocks in a *U-Net* are really important for correctly segmenting the objects, and which can be removed or skipped to save computation time without significant degradation of prediction quality. Our contribution is primarily a way of optimizing a neural network architecture. However, identifying the levels of detail at which the objects vary can also

This work was supported from the statutory research of Mechanical Faculty of WUST. The source material (images from laser cladding process) for the dataset preparation was supplied by National Centre for Research and Development - Project AMpHOra - Additive Manufacturing Processes and Hybrid Operations Research for Innovative Aircraft Technology Development – INNOLOT/I/6/NCBR/2013.

be seen as an important insight, helpful in better understanding of the problem.

## II. RELATED WORKS

The original U-Net [2] builds on the concepts of Fully Convolutional Networks [3]. While the FCN allowed using only some of the earlier layers to improve the fidelity of segmentation, U-Net's core concept is to merge even the most early blocks to capture high-resolution features. Further development on these ideas included Pyramid Scene Parsing [4] – where the input is sequentially pooled into separately processed streams and then upsampled and merged together before final prediction – and Feature Pyramid Networks [5], similar to U-Nets except that at every scale a complete segmentation is produced.

Optimization of neural network architectures was always of great interest. Early attempts such as Optimal Brain Surgeon [6] were primarily focused on improving the generalization capability of the learner. In more recent days, most architecture optimization work is focused on improving inference performance or energy efficiency [7], but there are also attempts to use these techniques to help extract classification rules [8]. The two major directions in network structure optimization are: architecture search and network pruning. The objective of architecture search is to find the optimal network structure during training, often using genetic algorithms or growing/pruning strategies [9, 10]. Network pruning focuses on removal of inactive or inefficient units from an already trained network in order to preserve its predictive power but reduce inference time [11].

There is not much research examining the influence of particular levels of detail on the object segmentation or detection quality. Chevalier *et al.* [12] studied the influence of input image resolution on classification performance, however they did not investigate the influence of deeper, highly downsampled layers. In this work, we propose a method of optimizing not only the size of the network input, but also its intermediate levels of detail as well.

## III. EXPERIMENT SETUP

### A. Scale-specialized blocks

U-Net consists of distinct “blocks”, comprising two 3x3 convolutional layers of various kernel depths, each followed by ReLU nonlinearity. From here onwards we will refer to them as simply *blocks*. Blocks are usually separated by max-pooling (in the “down” path) or upsampling and merge layers (in the “up” path). Thus, different blocks learn to extract features on different levels of detail.

Intuitively, depending on the characteristics of the problem, some of those blocks might be less useful for segmentation. This would mean that features of the data at these levels of detail are not important for a proper recognition. Blocks detecting those features would therefore waste compute power and memory. However, the problem of identifying them is not trivial.

Naively, one could envision training and comparison of multiple networks with different selection of blocks (e.g. one with 3 blocks and downsampling by a factor of 2, or 2 blocks and downsampling by 4). Such a brute-force ap-

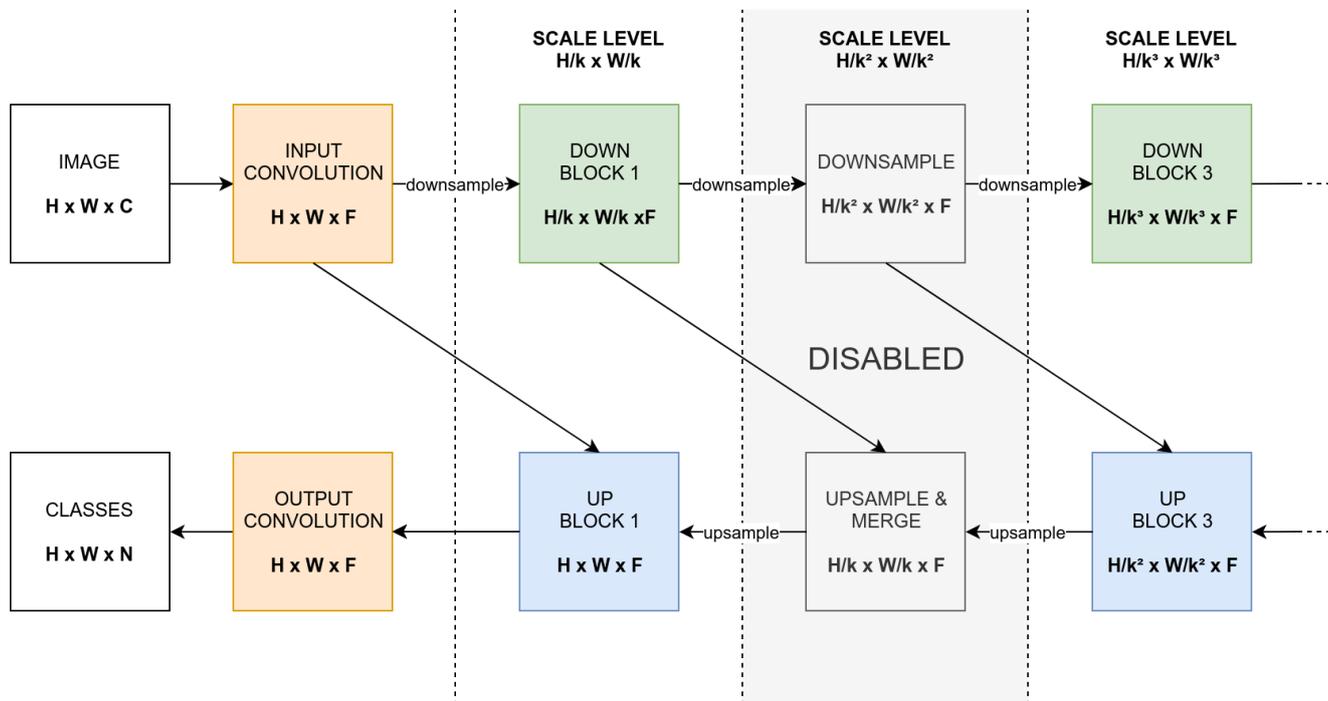


Fig 1. Drop-path regularization algorithm adapted to the general U-Net. In this example, “down” block 1 detects features from data downsampled to some specific resolution (*scale level*), and the corresponding “up” block integrates context extracted the corresponding level. Level 2 is shown in a disabled state – the “down” block only downsamples the data, while the “up” block only upsamples and merges it, both without any other processing.

proach might be infeasible, especially if the dataset is large and the network to be optimized is very deep. Ideally, a single network would be designed and trained in such a way that individual blocks could be freely removed from it without causing a structural failure, but instead only degrading performance – in the case the block was actually useful for prediction. Identifying the useless blocks would then proceed in a manner resembling a structural kind of ablation study.

### B. Drop-path regularization

Larsson *et al.* [13] presented a regularization algorithm, *drop-path*, that allowed them to train a very deep, multi-path network so that it behaves like an ensemble of networks. The core idea of drop-path is that if, during every training iteration, a random subset of individual paths in the network is disabled, the rest of the net will be forced to learn to still produce a correct answer. This allows the network to learn robustness against random removal of some sub-paths. Effectively, even though the network trains as a whole, every sub-path tries to become a fully capable standalone predictor itself. Larsson *et al.* report that they were able to extract even a single path of their FractalNet and it still worked almost as good as the whole.

We adapt the drop-path concept to U-Nets in order to allow them to learn robustness against removal of particular *levels of detail*. In a U-Net, the information from a particular scale is utilized twice during a single pass: once in the “down” path, where the features are extracted, and once again in the “up” path where the features are used to improve prediction resolution. Therefore, in our version of drop-path, whenever we randomly disable a “path”, we actually disable both blocks processing data on a particular scale. Overview of the algorithm is shown in Fig. 1.

To allow uninterrupted flow of the data through the disabled blocks, we replace each disabled “down” block with a simple bilinear downsampling layer, and the corresponding “up” block with a similar upsampling layer. We expect the network to learn to segment the images in the absence of information from particular scales, thus allowing evaluation of their influence on segmentation performance by means of a structural ablation test.

### C. Simplified U-Net

As the original architecture, U-Net does not naturally accommodate images of every size, requiring cropping and matching between “down” and “up” blocks, depending on the input size. However, as a meta-architecture it is very scalable – one can easily add or remove deeper blocks at different scales in order to capture more or less context in the data. We introduce several changes to the U-Net architecture to simplify it and make it more suitable for the drop-path regularization algorithm.

We add zero padding (1px wide border) to every convolutional layer, making each block preserve its input size. This eliminates the need for complex cropping and matching of data tensors throughout the “up” path.

We set every convolution in every block to produce the exact same number of channels (64), making every layer have exactly the same number of parameters. This is crucial in implementing drop-path: if different blocks produced outputs of different depths (as in the original U-Net), skipping a connection would necessitate a non-trivial mapping between the tensors.

Additionally, we introduce BatchNorm [14] after every convolution layer in order to stabilize the gradients. This is particularly important in the “up” path where data from two separate sources is combined.

Finally, following the practice of FPN, we change the type of connections between the “up” and “down” paths from concatenation (as originally in U-Net) to addition. This forms a residual connection between the paths, similar as described in [15]. This is not a critical change, but it reduces the number of parameters in the “up” convolutions by a factor of two, additionally speeding up the computation.

### D. Evaluation by ablation

We expect such a U-Net, trained using drop-path regularization, to behave like an ensemble of smaller networks, each processing data at a particular level of detail. This ensemble should be robust against removal of one member – at most, this should cause the overall performance to degrade, if that member (scale path) strongly contributes to the ensemble’s response. Therefore, we can measure the influence of a particular scale level by a structural ablation study. To test how important a particular level of detail is, we disable its corresponding block and evaluate the network on a validation set, measuring the change in segmentation performance. Additionally, we measure the average inference time to estimate the influence of disabling a block on wall-clock performance of the network.

In the experiment to follow, we use this evaluation strategy to reason about the data – and thus the problem at hand – in two ways.

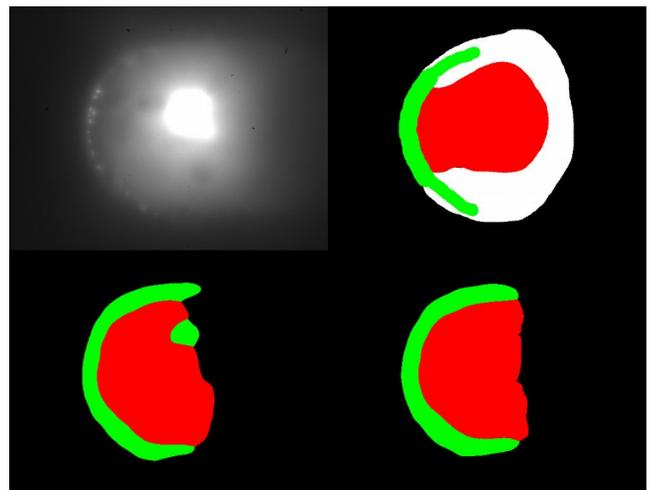


Fig 2. Example data and segmentations. Top row, left to right: source image, ground truth (“ignore” label in white); bottom row: segmentations – left: reduced model (see results, section B), right: full model.

By progressively disabling all blocks starting from the most high-resolution one, we attempt to identify the minimum scale level at which the network can observe the input images while still reliably segmenting the objects. The goal of this experiment is similar to Chevalier *et al.*, except we consider segmentation instead of classification.

By disabling subsequent blocks, starting from some given one, in different combinations, we attempt to find which of the intermediate levels of detail that extract contextual information are actually useful for a correct segmentation. This may provide an insight about how much context and on which level is really necessary, and which levels could be skipped to conserve compute time.

It is important to notice that the initial block of U-Net (on full scale) cannot be disabled – all subsequent layers require the input to be of a certain channel depth, and this first block transforms the original channel to a feature map of a common depth. This means that the initial convolutions will still be performed on the input of original resolution, constituting an approximately constant part of the computation time that cannot be trivially reduced.

#### IV. RESULTS

##### A. Reference network

We conduct the experiments on an in-house dataset of images acquired by coaxial on-line monitoring of a laser cladding process. Images obtained during this process are inherently noisy and blurry due to plasma emissions and powder scattering. However, they carry important information about process status, encoded in the shape of the pool of metal molten by the laser beam. The dataset consists of 250 grayscale images 600x600 pixels, manually annotated in 4 classes: background, two object classes of different shape characteristics (“edge” and “pool”) and an ignore label. Data was split in training and validation sets (150 and 100 im-

Parameter	Value
Learning rate schedule	constant 0.01
Adamax momenta	0.99, 0.999
Weight decay	0.0001
Batch size	64
Total iterations	750 000
Drop-path probability	0.25

ages, respectively). Example data and segmentations shown in Fig. 2.

The reference network consists of 5 levels of feature extraction blocks, at following scale levels: 600px, 300px, 150px, 75px, 25px and 5px. Each block comprises two 3x3 convolutional layers with 64 kernels, each followed by a BatchNorm layer and a ReLU nonlinearity. The network was trained using the Adamax [16] optimizer under the cross-entropy loss function. The complete training parameters are given in Table I.

Due to a small number of data samples and the need to train from scratch, heavy data augmentation routine was used in the form of elastic deformations [17] and horizontal and vertical flips. All augmentations were performed on-line in a random manner, directly before feeding data into the network. For testing, the intersection-over-union (IoU) metric was used. Results are given separately for either object class, due to different characteristics of their shapes.

Experiments were conducted in the PyTorch framework [18] using a single Nvidia RTX 2080 Ti GPU for training and an Nvidia TITAN Z for performance testing.

The reference network trained in approximately 11 hours achieving an IoU metric of 0.654 for the “edge” class and 0.809 for “pool” class. The average inference time (with gradient computation disabled) was 154.5ms, which is already approximately 38% faster than the original U-Net.

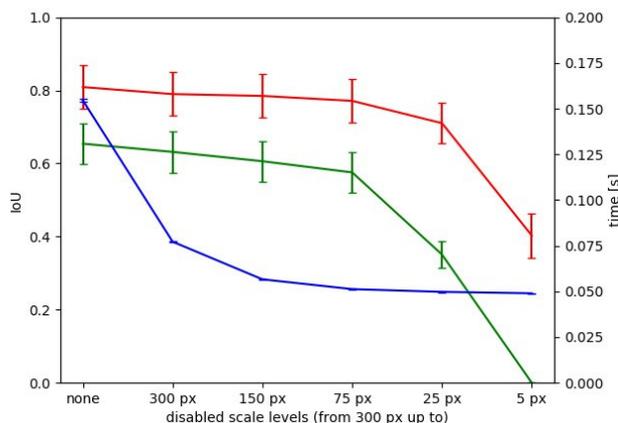


Fig 3. Results of the input size study. Segmentation performance (IoU for both classes, red and green plots) on the left axis, inference time (blue plot) on the right axis.

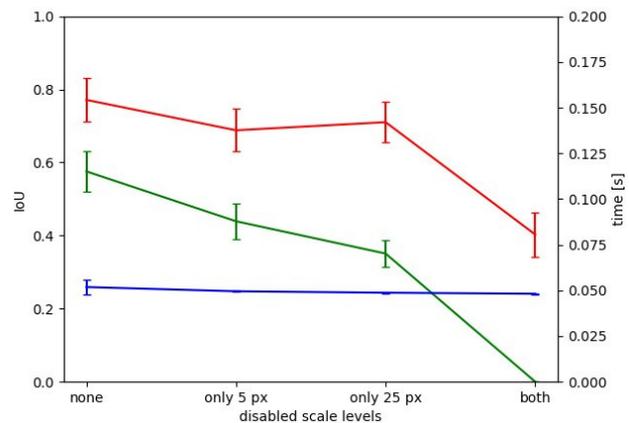


Fig 4. Results of the context levels study. Annotations the same as in Fig. 4.

### B. Input size study

Results of the progressive structural ablation study show relatively small degradation of segmentation quality when disabling the early, high-resolution blocks. As illustrated in Fig. 3, removing only the first block cuts the inference time in half, down to 77.2ms while only reducing the IoU score by 0.02. Performance improvements continue to be significant up until the scale levels of 150-75px, saturating at approximately 50ms (80% reduction with respect to the original U-Net). After that point, any potential speed-up is smaller than the cost of repeatedly downsampling the data to the desired resolution, while segmentation accuracy falls rapidly.

Fig. 2 visually compares segmentations produced by a full model (bottom right) and a model reduced by disabling scale levels 300 px and 150 px (bottom left).

### C. Context levels study

By disabling blocks at lower scale levels, we can determine the influence of particular context sizes. In this example we disable the first 2 levels (300px and 150px scales), fix the 75px scale as enabled and proceed to disable the remaining scale blocks (25px and 5px) in different combinations. We observe that in our case, disabling any level of context beyond the initial scale of 75px causes a rapid deterioration of segmentation quality, while providing zero practical improvement in inference time. Notice in Fig. 4 how disabling scale level 25px has a much more significant effect on the “edge” class (green) than on “pool” (red). Those most contextual blocks operate at very high relative scales (downsampling by the factors of 3 and 5, respectively) – we surmise that due to this, they learn very independent features that are critical to correct segmentation.

## V. CONCLUSION

In this work we presented a simplified and parameterized version of U-Net and adapted the drop-path algorithm to help the network learn as an ensemble of blocks specialized to detect features at specific levels of detail. This allowed us to analyze the importance of individual blocks on the collective network using a structural ablation study. That in turn let us identify blocks that did not contribute significantly to the segmentation, enabling us to make an informed decision to remove them in order to save compute time.

We argue that if the block was deemed unimportant, this might mean that at this particular scale there are no valuable features to be extracted – the data itself contains little valuable information. Therefore, processing inputs at this size is not worth the compute time. Aside from being a useful finding for optimizing a solution, this might also be a valuable insight into the nature of the problem itself.

### ACKNOWLEDGMENT

We wish to thank Mariusz Mrzygłód for valuable comments and discussions.

## REFERENCES

- [1] W. Rafajłowicz, P. Jurewicz, J. Reiner, and E. Rafajłowicz, “Iterative Learning of Optimal Control for Nonlinear Processes With Applications to Laser Additive Manufacturing,” *IEEE Transactions on Control Systems Technology*, pp. 1–8, 2018. <https://doi.org/10.1109/TCST.2018.2865444>
- [2] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 2015, pp. 234–241. [http://dx.doi.org/10.1007/978-3-319-24574-4\\_28](http://dx.doi.org/10.1007/978-3-319-24574-4_28)
- [3] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 3431–3440. <https://doi.org/10.1109/TPAMI.2016.2572683>
- [4] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “Pyramid Scene Parsing Network,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 6230–6239.
- [5] T. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature Pyramid Networks for Object Detection,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 936–944. <https://doi.org/10.1109/CVPR.2017.106>
- [6] B. Hassibi and D. G. Stork, “Second order derivatives for network pruning: Optimal brain surgeon,” in *Advances in neural information processing systems*, 1993, pp. 164–171.
- [7] S. Han, H. Mao, and W. J. Dally, “Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding,” presented at the *International Conference on Learning Representations (ICLR 2016)*, 2017
- [8] H. Lu, R. Setiono, and Huan Liu, “Effective data mining using neural networks,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 8, no. 6, pp. 957–961, 1996. <https://doi.org/10.1109/69.553163>
- [9] C. Liu et al., “Progressive Neural Architecture Search,” arXiv:1712.00559 [cs, stat], Dec. 2017. [preprint]
- [10] R. Luo, F. Tian, T. Qin, E. Chen, and T.-Y. Liu, “Neural Architecture Optimization,” in *Advances in Neural Information Processing Systems 31*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds. Curran Associates, Inc., 2018, pp. 7816–7827.
- [11] P. Molchanov, S. Tyree, T. Karras, T. Aila, and J. Kautz, “Pruning Convolutional Neural Networks for Resource Efficient Inference,” presented at the *International Conference on Learning Representations (ICLR 2017)*, 2017.
- [12] M. Chevalier, N. Thome, M. Cord, J. Fournier, G. Henaff, and E. Dusch, “Low resolution convolutional neural network for automatic target recognition,” in *7th International Symposium on Optronics in Defence and Security*, Paris, France, 2016.
- [13] G. Larsson, M. Maire, and G. Shakhnarovich, “FractalNet: Ultra-Deep Neural Networks without Residuals,” arXiv:1605.07648 [cs], May 2016. [preprint]
- [14] S. Ioffe and C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” in *International Conference on Machine Learning*, 2015, pp. 448–456.
- [15] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778. <https://doi.org/10.1109/CVPR.2016.90>
- [16] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” in *3rd International Conference on Learning Representations, ICLR 2015*, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings, 2015.
- [17] P. Y. Simard, D. Steinkraus, and J. C. Platt, “Best practices for convolutional neural networks applied to visual document analysis,” in *Seventh International Conference on Document Analysis and Recognition*, 2003. Proceedings., 2003, pp. 958–963.
- [18] A. Paszke et al., “Automatic differentiation in PyTorch,” in *NIPS-W*, 2017.



# Predicting Blood Glucose using an LSTM Neural Network

Touria El Idriss  
Department of Computer Sciences  
EMI, University Mohamed V  
Rabat, Morocco  
Email: {el.idrissit@gmail.com}

Ali Idri and Ibtissam Abnane  
Software Project Management Research  
Team ENSIAS, University Mohamed V  
Rabat, Morocco  
Email: {ali.idri@um5.ac.ma,  
ibtissam.abnane19@gmail.com}

Zohra Bakkoury  
Department of Computer  
Sciences EMI, University  
Mohamed V Rabat, Morocco  
Email: {bakkoury@gmail.com}

**Abstract**—Diabetes self-management relies on the blood glucose prediction as it allows taking suitable actions to prevent low or high blood glucose level. In this paper, we propose a deep learning neural network (NN) model for blood glucose prediction. It is a sequential one using a Long-Short-Term Memory (LSTM) layer with two fully connected layers. Several experiments were carried out over data of 10 diabetic patients to decide on the model's parameters in order to identify the best variant of it. The performance of the proposed LSTM NN measured in terms of root mean square error (RMSE) was compared with the ones of an existing LSTM and an autoregressive (AR) models. The results show that our LSTM NN is significantly more accurate; in fact, it outperforms the existing LSTM model for all patients and outperforms the AR model in 9 over 10 patients, besides, the performance differences were assessed by the Wilcoxon statistical test. Furthermore, the mean of the RMSE of our model was 12.38 mg/dl while it was 28.84 mg/dl and 50.69 mg/dl for AR and the existing LSTM respectively.

## I. INTRODUCTION

DATA mining (DM) techniques are useful tools for extracting valuable knowledge from (large) databases that helps in decision making [1], [2]. DM has been fruitfully used in different subfields of medical informatics such as diabetes [1], [3], cardiology [1], [4] and cancer [1], [5].

This paper deals with the application of DM for diabetes which is a chronic illness caused by a disorder in the glucose metabolism. There are mainly two types of diabetes): 1) Type 1 Diabetes Mellitus (T1DM) when the pancreas does not produce enough insulin, and Type 2 Diabetes Mellitus (T2DM) which results from an ineffective use of insulin [6]. If not well managed, this disease can lead to serious problems such as heart attacks, kidney damage, blindness, unconsciousness and even death [6]. The prediction of blood glucose level (BGL) is an important task in the diabetes management and self-management as it can help controlling the BGL by taking appropriate actions ahead of time [7]. To

predict the BGL, the previous glucose measurements are required. The BGL can be measured: 1) Manually by self-monitoring of blood glucose (SMBG) using sticks several times a day or 2) Automatically by continuous glucose monitoring (CGM) using sensors [6], [7].

According to El Idrissi et al. [7], considerable work was done for the BGL prediction and various Data Mining approaches including statistical methods and machine learning techniques were investigated for that purpose; the most used ones are Artificial Neural Networks (NNs) and Auto Regression (AR) [7]. Recently, deep learning modeling is gaining more interest, such as LSTM NN [8] and deep NN [9].

This paper proposes a deep learning NN with one LSTM layer and two fully connected layers for the prediction of BGL using CGM data. Predicting glucose using LSTM Nns is promising [8] since LSTM NNs were successfully applied in other domains such as prediction of water quality [10], electricity consumption [11] and stock prices [12].

This work aims at: (1) Setting the parameters of the model to identify the best configuration of our LSTM NN; and (2) Assessing and comparing the accuracy of the proposed model to existing ones. Toward this aim, two research questions are discussed:

(RQ1): Can the proposed LSTM model achieve good performance?

(RQ2): Is the proposed LSTM model significantly more accurate than existing models?

This paper is structured as follows: Section II presents an overview of LSTM NNs. Section III summarizes the related work on predicting blood glucose. Data used and performance measurement are described in Section IV. Section V describes the experimental design. Results are reported and discussed in Section VI. Threats to validity of this study are presented in Section VII and finally conclusion and future work are presented in Section VIII.

## II. LSTM NNS: AN OVERVIEW

Deep learning in NNs is an emerging method that allows the NN to learn automatically the characteristics of data by selecting the relevant features [10], contrary to the classical NNs that require features' selection based on domain knowledge [9].

LSTM NNs are deep recurrent NNs (RNNs) that were introduced by Hochreiter and Schmidhuber [13] to overcome the problem of exploding or vanishing gradient encountered with traditional RNNs [8]. The LSTM NNs are suitable for sequential data such as speech, video and time series as they can capture long term dependencies [14]. They consist of memory cells with a cell state which is maintained over time, and a gate structure that controls and regulates the information of the cell state.

Fig. 1 illustrates the structure of a memory cell. The index  $t$  refers to time or sequence.  $X_t$ ,  $Y_t$ ,  $h_t$  and  $C_t$  represent respectively the input, the output, the hidden vector and the cell state for  $t$ .

The memory cell contains 3 gates: 1) Input gate selects the information to be retained in the cell; 2) Forget gate decides about the information to be ignored; and 3) Output gate calculates the output and updates the hidden vector. Each of these gates is a NN whose input vector is a concatenation of the hidden vector of the previous cell and the input vector. Let  $W_i$ ,  $W_f$ ,  $W_o$  be the weight matrices corresponding respectively to the input, forget and output gates; and  $b_i$ ,  $b_f$ ,  $b_o$  the corresponding bias vector.  $W_c$  and  $b_c$  are the weight matrix and the bias vector used for updating the cell state.

The result of the input gate is  $i_t$ , which is obtained as follows:

$$i_t = \sigma(W_i * [h_{t-1}, X_t] + b_i) \quad (1)$$

To calculate  $f_t$ , the forget gate uses the following equation:

$$f_t = \sigma(W_f * [h_{t-1}, X_t] + b_f) \quad (2)$$

The output gate uses the equation (3) to obtain  $o_t$ , and the equation (4) to get the hidden vector.

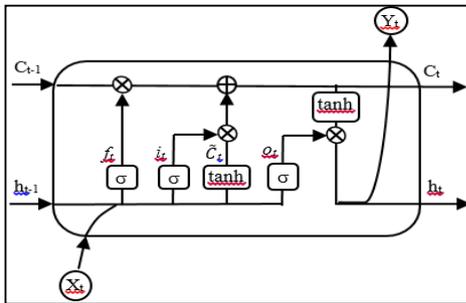


Fig. 1 : Memory cell

$$o_t = \sigma(W_o * [h_{t-1}, X_t] + b_o) \quad (3)$$

$$h_t = o_t * \tanh(C_t) \quad (4)$$

The cell state is updated as follows:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (5)$$

$$\text{where } \tilde{C}_t = \tanh(W_c * [h_{t-1}, X_t] + b_c) \quad (6)$$

In the Equations (1 to 4 and 6),  $\sigma$  and  $\tanh$  are the activation functions, the former is the sigmoid function defined in Equation (7) and the latter is the hyperbolic tangent function defined in equation (8).

$$\sigma(x) = \frac{e^x}{1+e^x} \quad (7)$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (8)$$

$\sigma$  and  $\tanh$  are defined on the set of real numbers.  $\sigma$  ranges between 0 and 1 while  $\tanh$  ranges from -1 to 1.

## III. RELATED WORK

El Idrissi et al. [7] investigated by the mean of a systematic map and review the use of predictive techniques in data mining for the diabetes self-management. The study summarized and reviewed 38 studies published from 2000 to April 2017 with regards to: 1) publication's year and source, 2) diabetes' type, 3) investigated clinical tasks, 4) the used predictive techniques and 5) their performance.

The main findings of [7] are the following:

1. There is a growing interest to the use of DM predictive techniques in the last decade. Conferences and Journals are used for publication, nevertheless, the main publication channel is conferences by 76.32% of the selected studies, while just 23.68% were published in journals.

2. T1DM gained more attention for research than T2DM with 84.21% of the selected studies.

3. Considerable work was done for the BGL prediction comparatively to other clinical tasks, in fact, 57.89% of the considered papers investigated BGL prediction.

4. Various DM predictive techniques were investigated, and the most used ones are NNs and AR.

5. NNs and AR models yield the highest accuracies. However, none of the used DM predictive techniques is dominant over the others.

Table I reports the findings of a set of selected studies from the review of El Idrissi et al. [7], in addition to the two recent studies [8], [9]. From Table I, we note that:

1. Various techniques are investigated to the BGL prediction problem: statistical methods such as AR [15], [16] and Kalman Filter (KF) [17]; machine learning methods such as Artificial NNs [18]-[20] and Support Vector Regression (SVR) [21], [22]; and recently deep learning techniques [8], [9].

TABLE I.  
LITERATURE OVERVIEW OF THE BGL PREDICTION

Reference	Technique	Diabetes type	Input Data	Findings
Lu et al., 2010 [15]	AR	T1DM	CGM	AR models yield accurate BGL prediction with short length signals; it is not required to consider exogenous inputs explicitly nor all frequency bands of the glucose signals.
Novara et al., 2016 [16]	AR	T1DM	CGM	A blind identification using AR technique was proposed to predict the BGL and recover the unmeasured inputs.
Wang et al., 2013 [17]	KF	T1DM	CGM	The model based on an extended KF achieves mostly reliable BGL predictions, and made significant improvement compared to zero-order hold.
Zarkogianni et al. 2011 [18]	RNNs	T1DM	CGM	The proposed RNN model to simulate the blood glucose–insulin metabolism makes it possible to personalize the system and to handle the environment’s variations with efficiency.
Allam et al., 2011 [19]	Feedforward NNs	T1DM	CGM	For short prediction horizon, feedforward NN based model gets accurate BGL prediction without time lagging.
Mathiyazhagan & Schechter, 2014 [20]	Fuzzy NNs	T1DM	CGM	This study proposes a soft computing approach which tolerates imprecision by using a fuzzy NN to predict the BGL.
Bunescu et al., 2013 [21]	SVR	T1DM	CGM	The incorporation of physiological features into an existing SVR model for BGL prediction made a significant performance enhancement .
Georga et al., 2010 [22]	SVR	T1DM	CGM	Predicting BGL is possible by compartmental models and SVR with a satisfactory accuracy and a clinical acceptability.
Sun et al, 2018 [8]	LSTM NN	T1DM	CGM	An LSTM network with one LSTM layer followed by one bi-directional LSTM layer and several fully connected layers were proposed for BGL prediction. This LSTM model outperformed the baseline methods ARIMA and SVR.
Mhaskar et al, 2017 [9]	Deep NN	T1DM	CGM	The proposed deep NN for BGL prediction outperforms shallow NN.

2. Different types of Artificial NNs were explored: RNNs [18], feedforward NNs [19], fuzzy NNs [20], and deep NNs [8], [9].

3. Studies are using data coming from CGM for T1DM patients.

#### IV. DATA DESCRIPTION AND PERFORMANCE MEASUREMENT

The data set used in this study is the historical data set DirecNetInpatientAccuracyStudy provided by Diabetes Research in Children Network (DirecNet) [23]. It holds collected data of 110 T1DM patients particularly the recorded data from CGM devices which give the BGLs at intervals of 5 minutes. As we did not specify any prerequisite for our model, we considered randomly a subset of 10 patients. However, a pre-processing of the data was done to eliminate redundancy and outliers between consecutive measurements. Table II summarizes information about the considered patients.

To evaluate the performance of our model, we use the root mean square error (RMSE) which is a common performance metric used to assess BGL prediction [7]. It evaluates the difference between the actual value and the predicted one by means of the Equation (9):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{e}_i - e_i)^2} \quad (9)$$

where  $e_i$  and  $\hat{e}_i$  are respectively the actual and the predicted values while  $n$  is the sample’s size. The value of RMSE goes

from 0 to  $+\infty$  and the prediction is better when the RMSE is low.

#### V. EXPERIMENTAL DESIGN

This paper proposes an LSTM NN to predict the BGL of diabetic patients. This NN uses one LSTM layer and two dense layers. To identify the best configuration to adopt, we performed by means of a Grid Search (GS) described in Table III and inspired from [24], the following steps:

Step 1: Train and test the model for each patient by varying the number of LSTM units (LU) according to the GS values of Table III. The chosen value is the one that gives the best RMSE over the patients’ datasets.

TABLE II.  
DATA SET DESCRIPTION. THE BGL IS IN MG/DL

Patient ID	Number of measurements	Min BGL	Max BGL	Mean BGL
PT01	766	40	339	114.78
PT02	278	57	283	120.96
PT03	283	103	322	185.89
PT04	923	40	400	188.44
PT05	562	50	270	179.71
PT06	771	62	400	187.45
PT07	897	42	400	210.26
PT08	546	43	310	152.88
PT09	831	40	400	157.50
PT10	246	72	189	116.51

Step 2: Using the LU obtained in Step 1, we train and evaluate the model for each patient by varying the number of dense units (DU) respecting the GS values of Table III. The chosen value is the one that gives the best RMSE over the patients' datasets.

Step 3: Using the LU and DU obtained in Step 1 and Step 2 respectively, we train and evaluate the model for each patient by varying the length of the input sequences (SL) according to the GS values of Table III. The chosen value is the one that gives the best RMSE over the patients' datasets.

Step 4: We compare in terms of the RMSE criterion our best LSTM model with the LSTM of [8] and an AR model.

Step 5: We evaluate the statistical significance of performance differences between the three models by means of the Wilcoxon statistical test [25].

## VI. RESULTS AND DISCUSSION

This section presents and discusses the obtained results: (1) We present the empirical results of the Steps 1 to 3 of the experimental process related to the parameters' setting. (2) The results of the performance comparison of our model to the two other models are presented. (3) We report the results of the statistical test. And (4), We discuss all the empirical results.

Our LSTM model is developed using Python 3.6 with the framework Keras 2.2.4 and Tensorflow 1.12.0 as backend and under the operating system Windows 10.

### A. Parameters selection

The main objective of this research is to propose an LSTM model that achieves good performance (RQ1). To do that, a phase of parameters selection was carried out. This phase consists of the three first steps of the experimental design of Section V. The results of the first step are shown in Fig. 2 which represents the RMSE for each patient according to the search space of LU defined in Table III. We observe that the RMSE is in general better for LU=50. Therefore, at that step the LU was chosen to be 50 units.

TABLE III.  
PARAMETERS FOR SEARCH GRID

Parameter	Signification	Search space
LSTM Units	Number of neurons in the gates	{5, 10, 20, 30, 40, 50, 60, 70}
Dense Unit	Number of the neurons in the dense layer	{10, 20, 30, 40, 50}
Sequence length	Dimension of the input vector	{5, 10, 15, 20}

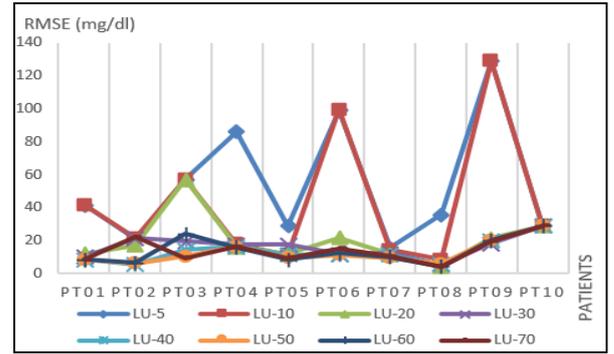


Fig. 2 : RMSE with LU variations

In step 2, we did the same experimentations with LU=50 and we vary DU according to the GS values of Table III. Fig. 3 shows the results obtained: we observe that for DU=40 and DU=50, the RMSE is worse than the others. Furthermore, the values 10, 20 and 30 of DU gave comparable results. However, by considering the times that a configuration (DU is equal to 30, 20 or 10) is better than the others, the minimum RMSE was reached 4 times, 3 times and once respectively. Therefore, the chosen DU value is 30.

The Step 3 uses LU=50, DU=30, and varies the SL according to the GS values of Table III. Fig. 4 shows that SL=10 achieves in general better RMSE values.

To sum up, the best configuration of our LSTM uses LU=50, DU=30 and SL=10. This LSTM variant will be used to compare it with existing BGL predictors.

### B. Models Comparison

To answer RQ2, we applied the LSTM model of [8] referred to as Sun\_LSTM and AR on the same dataset. We carried out the comparison with Sun\_LSTM since it is, according to the best of our knowledge, the only study that proposed an LSTM model for BGL prediction. Regarding the comparison with an AR model, it is motivated by the fact that AR models are suitable for time series prediction and according to [7] they yield along with NNs highest accuracy rates.

The Sun\_LSTM contains one LSTM layer with 4 LU followed by one bi-directional LSTM layer with 4 LU and 3 fully connected dense layers with respectively 8, 64 and 8 DU and finally the output Dense Layer. The SL is set to 4. Regarding the AR model, we used the default model developed in Weka with a lag of 10 similarly to the SL proposed in our model which is equal to 10.

Fig. 5 presents the RMSE got for the three models: we observe that our LSTM model outperforms Sun\_LSTM for all patients and outperforms AR in 9 cases over 10 patients.

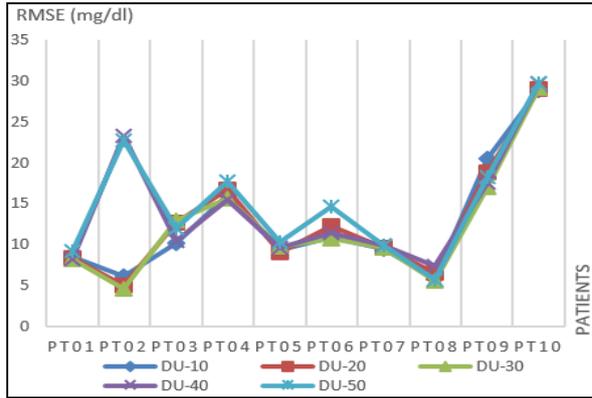


Fig. 3: RMSE with DU variations

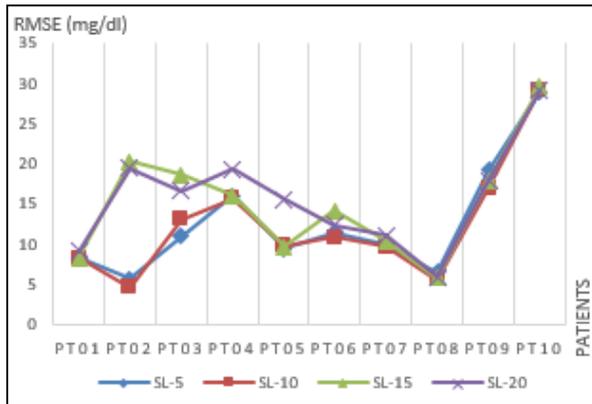


Fig. 4: RMSE with SL variations

Furthermore, we have compared the training time for the two LSTM models presented in Fig. 6. Our model needs overall less time to be trained comparatively to Sun\_LSTM.

Finally, we recorded the number of epochs required for each patient to train the model. Results are presented in Table IV.

The number of epochs is low and ranges between 46 and 260, the mean is equal to 132.8.

TABLE IV.  
NUMBER OF EPOCHS

Patient ID	Number of epochs	Patient ID	Number of epochs
PT01	91	PT06	260
PT02	94	PT07	46
PT03	259	PT08	49
PT04	71	PT09	168
PT05	88	PT10	81

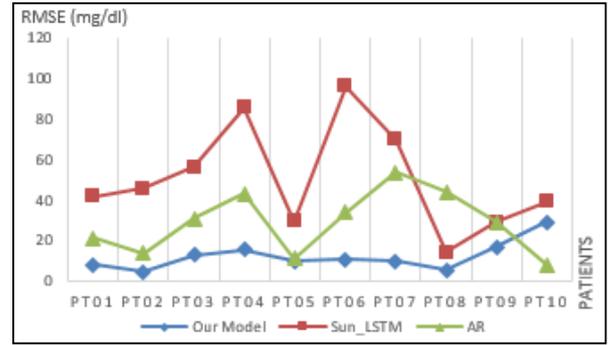


Fig. 5: RMSE for Our model, Sun\_LSTM and AR

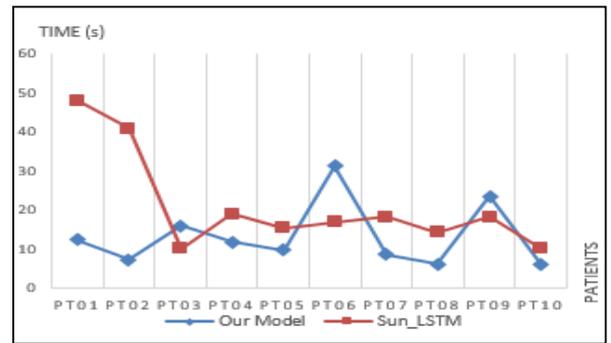


Fig. 6: Training time for Our model and Sun\_LSTM

C. Statistical test

We statistically assessed the results obtained based on the Wilcoxon statistical test. It is a non-parametric test that validates if the differences between the compared models are statistically significant by the means of a statistical hypothesis [25]. In our case, two null hypotheses (NH) were considered to test if our LSTM model is better than Sun\_LSTM and AR. These hypotheses are the following:

- NH1 : Our LSTM model does not outperform Sun\_LSTM.
- NH2 : Our LSTM model does not outperform AR model.

The statistical test was two-tailed considering the significance level  $\alpha$  equals to 0.05. Note that the test is considered significant if the result which is the p-value is less than  $\alpha$ .

For our tests, the p-value of NH1 and NH2 were 0.00512 and 0.0285 which means that the performance differences between our models and respectively the two other models were statistically significant.

D. Discussion

In this study, we propose an LSTM NN to predict BGL. Our first concern was to determine the best configuration by tuning the three following parameters: LU, DU and SL. As we observe from the results of Fig. 2, Fig. 3 and Fig. 4, these parameters have an impact on the model's performance. Thus, when building an LSTM model, one has to determine

the optimal values of these parameters. The final model achieved good accuracy comparatively to what was found in literature by [7]. In fact, for the RMSE, the minimum, maximum and mean values were equal to 4.67, 29.12 and 12.38 respectively. On the other hand, our model outperforms a previous LSTM [8] and AR models. Furthermore, the performance of our model over the two others was assessed by the Wilcoxon statistical test which showed that our model significantly outperformed Sun\_LSTM and AR.

These results show that the use of LSTM NNs for BGL prediction is promising. In fact, LSTM NNs and Deep NNs in general are gaining more interest due to their performances in different fields [26]. Actually, the deep NNs took advantage from their training algorithms that are computationally efficient [27], and from the large hidden neurons' number which results on huge number of free parameters [26]. On another hand, the performance of our model over Sun\_LSTM can be explained by the fact that our LSTM is deeper than Sun\_LSTM: in fact, the former uses 50 LU while the latter uses 4 LU.

However, we believe that the accuracy can be improved by considering: 1) other parameters to tune such as learning rate, the loss function and the optimizer, and 2) more information as input such as medication and exercise.

From another point of view, the training time and the epoch numbers were low, which means that the model can be used on a mobile platform or wearable device [7] which require rapidity [8]. In fact, knowing the importance of the mobile personal health records in health management [28], [29], the deployment of our model on a mobile platform is another direction to explore. Toward that aim, we used TensorFlow as it provides an end-to-end support which goes till the deployment of the models on mobile apps, cloud server and others [30].

#### VII. THREADS TO VALIDITY

Four of threads to validity are identified which are:

**Internal validity:** This thread concerns the evaluation methodology which can be inappropriate. To overcome this limitation, we used 10 data sets, each of them was divided in training and test sets representing respectively 66% and 34% of the whole data set. The model was trained on the former and evaluated on the latter.

**External validity:** This thread is related to the perimeter of validity. This was settled by considering 10 diabetic patients randomly taken from a public dataset. Note that some studies used only one dataset as reported in [7]. Furthermore, the datasets are with different sizes, it goes from 246 to 923 instances.

**Construct validity:** It concerns the measurement validity. To limit this thread, we used the RMSE which is a common performance criterion used to assess BGL prediction [7].

Moreover, it was used by many studies such as [8], [15], [18], [19], [21], [22]. Note that it was used lonely by [15], [21].

**Statistical conclusion validity:** It affects the conclusion related to the comparison performed. To avoid this thread, we used the Wilcoxon statistical test to assess the significance of the performance differences.

#### VIII. CONCLUSION AND FUTURE WORK

This paper considered a deep NN for BGL prediction which is a sequential model with one LSTM layer and two fully connected layers. Multiple runs were done by varying the three parameters: LU, DU and SL to determine the best configuration. The model achieves good accuracy and significantly outperforms, based on RMSE, an existing LSTM model and an AR model.

These promising results encourage to carry out further investigations using the LSTM NNs. Ongoing work aims at building an LSTM model for multi-step prediction by exploring different strategies from literature. Considering more input data such as medication and exercise may improve the accuracy. Furthermore, a problem that can be encountered is the missing data when for example the patient takes off the CGM device, investigating how to handle automatically the missing data is an interesting direction to consider.

#### REFERENCES

- [1] N. Esfandiari, M. R. Babavalian, A. M. E. Moghadam & V. K. Tabar, "Knowledge discovery in medicine: Current issue and future trend", *In Expert Systems with Applications*, vol. 41, no. 9, pp. 4434-4463, 2014, <https://doi.org/10.1016/j.eswa.2014.01.011>
- [2] H. Benhar, A. Idri and J.-L. Fernández-Alemán, "Data preprocessing for decision making in medical informatics: potential and analysis." *World Conference on Information Systems and Technologies*. Springer, Cham, 2018, pp. 1208-1218, [https://doi.org/10.1007/978-3-319-77712-2\\_116](https://doi.org/10.1007/978-3-319-77712-2_116).
- [3] T. El Idrissi, A. Idri, and Z. Bakkoury, "Data Mining Techniques in Diabetes Self-management: A Systematic Map", *In World Conference on Information Systems and Technologies*, Springer, Cham, 2018, pp. 1142-1152, [https://doi.org/10.1007/978-3-319-77712-2\\_109](https://doi.org/10.1007/978-3-319-77712-2_109).
- [4] I. Kadi, A. Idri and J.-L. Fernandez-Aleman. "Knowledge discovery in cardiology: A systematic literature review". *International Journal of Medical Informatics*, vol. 97, pp. 12-32, 2017, <https://doi.org/10.1016/j.ijmedinf.2016.09.005>.
- [5] A. Idri, I. Chlioui and B. EL Ouassif, "A systematic map of data analytics in breast cancer", *In Proceedings of the Australasian Computer Science Week Multiconference*, ACM, 2018, p. 26, <https://doi.org/10.1145/3167918.3167930>.
- [6] R. Billous, R. Donnally, "Handbook of Diabetes", Blackwell, 2010.
- [7] T. EL Idrissi, A. Idri and Z. Bakkoury, "Systematic map and review of predictive techniques in diabetes self-management", *International Journal of Information Management*, vol. 46, pp. 263-277, 2019, <https://doi.org/10.1016/j.ijinfomgt.2018.09.011>.
- [8] Q. Sun, M. V. Jankovic, L. Bally and S. G. Mougiakakou, "Predicting Blood Glucose with an LSTM and Bi-LSTM Based Deep Neural Network," *In Symposium on Neural Networks and Applications (NEUREL)*, Belgrade, 2018, pp. 1-5, <https://doi.org/10.1109/NEUREL.2018.8586990>
- [9] H. N. Mhaskar, S. V. Pereverzyev, and M. D. van der Walt, "A Deep Learning Approach to Diabetic Blood Glucose Prediction," *Front. Appl. Math. Stat.*, vol. 3, no. July, pp. 1-11, 2017, <https://doi.org/10.3389/fams.2017.00014>

- [10] Y. Wang, J. Zhou, K. Chen, Y. Wang and L. Liu, "Water quality prediction method based on LSTM neural network," 2017 12<sup>th</sup> International Conference on Intelligent Systems and Knowledge Engineering (ISKE), Nanjing, 2017, pp. 1-5, <https://doi.org/10.1109/ISKE.2017.8258814>
- [11] N. Kim, M. Kim and J. K. Choi, "LSTM Based Short-term Electricity Consumption Forecast with Daily Load Profile Sequences," 2018 IEEE 7th Global Conference on Consumer Electronics (GCCE), Nara, 2018, pp. 136-137, <https://doi.org/10.1109/GCCE.2018.8574484>
- [12] M. Roondiwala, H. Patel, and S. Varma, "Predicting Stock Prices Using LSTM," International Journal of Science and Research (IJSR), vol. 6, no. 4, pp.1754-1756, 2017.
- [13] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Comput., vol. 9, no. 8, pp. 1735-1780, 1997, <https://doi.org/10.1162/neco.1997.9.8.1735>
- [14] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink and J. Schmidhuber, "LSTM: A Search Space Odyssey," in IEEE Transactions on Neural Networks and Learning Systems, vol. 28, no. 10, pp. 2222-2232, Oct. 2017, <https://doi.org/10.1109/TNNLS.2016.2582924>
- [15] Y. Lu, A. V. Gribok, W. K. Ward, and J. Reifman.. "The Importance of Different Frequency Bands in Predicting Subcutaneous Glucose Concentration in Type 1 Diabetic Patients," in IEEE Transactions on Biomedical Engineering, vol. 57, no. 8, pp. 1839-1846, Aug. 2010, <https://doi.org/10.1109/TBME.2010.2047504>
- [16] C. Novara, N. M. Pour, T. Vincent, G. Grassi, "A Nonlinear Blind Identification Approach to Modeling of Diabetic Patients," in IEEE Transactions on Control Systems Technology, vol. 24, no. 3, pp. 1092-1100, May 2016, <https://doi.org/10.1109/TCST.2015.2462734>
- [17] Q. Wang S. Harsh, P. Molenaar, and K. Freeman, "Developing personalized empirical models for Type-I diabetes: An extended Kalman filter approach," American Control Conference, IEEE, 2013, pp. 2923-2928, <https://doi.org/10.1109/ACC.2013.6580278>.
- [18] K. Zarkogianni, A.Vazeou, S. G. Mougiakakou, A. Prountzou and K. S. Nikita, "An Insulin Infusion Advisory System Based on Autotuning Nonlinear Model-Predictive Control," in IEEE Transactions on Biomedical Engineering, vol. 58, no. 9, pp. 2467-2477, Sept. 2011, <https://doi.org/10.1109/TBME.2011.2157823>.
- [19] F. Allam, Z. Nossair, H. Gomma, I. Ibrahim, and M. Abd-el Salam. "Prediction of subcutaneous glucose concentration for type-1 diabetic patients using a feed forward neural network," The Int. Conf. On Computer Engineering & Systems, Cairo, 2011, pp. 129-133, <https://doi.org/10.1109/ICCES.2011.6141026>.
- [20] N. Mathiyazhagan, H. B. Schechter, "Soft computing approach for predictive blood glucose management using a fuzzy neural network," IEEE Conf. on Norbert Wiener in the 21st Century (21CW), Boston, MA, 2014, pp. 1-3, <https://doi.org/10.1109/NORBERT.2014.6893906>.
- [21] R. Bunescu, N. Struble, C. Marling, J. Shubrook, and F. Schwartz, "Blood Glucose Level Prediction Using Physiological Models and Support Vector Regression," In 2013 12th International Conference on Machine Learning and Applications, vol. 1, pp. 135-140. IEEE, 2013, <https://doi.org/10.1109/ICMLA.2013.30>.
- [22] E. I. Georga, V. C. Protopappas, D. Polyzos, "Prediction of glucose concentration in type 1 diabetic patients using support vector regression," Proceedings of the 10th IEEE Int. Conf. on Information Technology and Applications in Biomedicine, Corfu, 2010, pp. 1-4, <https://doi.org/10.1109/ITAB.2010.5687764>
- [23] Diabetes Research in Children Network (DirecNet). Available online at: <http://direcnet.jaeb.org/Studies.aspx> [Ap. 1,2019]
- [24] M. Hosni, A. Idri and A. Abran, "Investigating heterogeneous ensembles with filter feature selection for software effort estimation." In Proceedings of the 27th International Workshop on Software Measurement and 12th International Conference on Software Process and Product Measurement, ACM, 2017, pp. 207-220, <https://doi.org/10.1145/3143434.3143456>
- [25] A. Idri, I. Abnane and A. Abran. "Missing data techniques in analogy-based software development effort estimation." Journal of Systems and Software, vol. 117, pp. 595-611, 2016, <https://doi.org/10.1016/j.jss.2016.04.058>.
- [26] X. Chen and X. Lin, "Big Data Deep Learning: Challenges and Perspectives," in IEEE Access, vol. 2, pp. 514-525, 2014, <https://doi.org/10.1109/ACCESS.2014.2325029>.
- [27] N. Papernot, P. McDaniel, S. Jha, M. Fredrikson, Z. B. Celik and A. Swami, "The Limitations of Deep Learning in Adversarial Settings," 2016 IEEE European Symposium on Security and Privacy (EuroS&P), Saarbrücken, 2016, pp. 372-387, <https://doi.org/10.1109/EuroSP.2016.36>.
- [28] S. Ouhbi, A. Idri, J. L. Fernández-Alemán and A. Toval, "Mobile personal health records for cardiovascular patients," 2015 Third World Conference on Complex Systems (WCCS), Marrakech, 2015, pp. 1-6, <https://doi.org/10.1109/ICoCS.2015.7483226>
- [29] M. Bachiri, A. Idri, J.-L. Fernández-Alemán, A. Toval. "Mobile personal health records for pregnancy monitoring functionalities: Analysis and potential." Computer methods and programs in biomedicine, vol. 134, pp. 121-135, 2016, <https://doi.org/10.1016/j.cmpb.2016.06.008>.
- [30] Sergeev, A., & Del Balso, M. (2018). Horovod: fast and easy distributed deep learning in TensorFlow. arXiv preprint arXiv:1802.05799.



# Accurate Retrieval of Corporate Reputation from Online Media Using Machine Learning

Achim Klein  
Information Systems 2, University  
of Hohenheim, 70599 Stuttgart,  
Germany  
Email: achim.klein@uni-  
hohenheim.de

Martin Riekert  
Information Systems 2, University  
of Hohenheim, 70599 Stuttgart,  
Germany  
Email: martin.riekert@uni-  
hohenheim.de

Velizar Dinev  
Information Systems 2, University  
of Hohenheim, 70599 Stuttgart,  
Germany  
Email: velizar.dinev@gmail.com

**Abstract**—Corporate reputation is an economic asset and its accurate measurement is of increasing interest in practice and science. This measurement task is difficult because reputation depends on numerous factors and stakeholders. Traditional measurement approaches have focused on human ratings and surveys, which are costly, can be conducted only infrequently and emphasize financial aspects of a corporation. Nowadays, online media with comments related to products, services, and corporations provides an abundant source for measuring reputation more comprehensively. Against this backdrop, we propose an information retrieval approach to automatically collect reputation-related text content from online media and analyze this content by machine learning-based sentiment analysis. We contribute an ontology for identifying corporations and a unique dataset of online media texts labelled by corporations' reputation. Our approach achieves an overall accuracy of 84.4%. Our results help corporations to quickly identify their reputation from online media at low cost.

## I. INTRODUCTION

A great variety of firms offer an even greater variety of products and services to consumers and other businesses and strive to build up a strong corporate reputation. Corporate reputation can be defined as the collective perception and judgment of the sentiment (i.e., feeling, opinion) about a corporation and its products or services by its stakeholders. Reputation as a necessary condition for differentiation and corporate success has become one of the central themes in all its facets for both practitioners and the scientific community [1]. The ability to quickly assess current movements in the own and the competitors' corporate reputation is crucial for operative decision making, corporate planning and strategy as well as for external investment decisions.

The important role of corporate reputation has been confirmed through extensive research. Shefrin and Statman show that corporations with good reputation represent good long-term investment opportunities [2]. These corporations with a good corporate reputation are more likely to receive funding on the capital markets at better conditions. The positive relationship between corporate reputation and investor expectations about a firm has been supported again later by Shefrin [3] and MacGregor et al. [4] pointing

toward a stable relationship. Corporate reputation is shown to be positively related to return on sales and assets, sales, earnings per share, price-to earnings ratio, dividend yield, net income of a company, and customer loyalty [5-8].

Studies in the field of corporate reputation [5-8] have in common that authors use either the Fortune magazine's reputation index published in the annual survey "Most admired companies" or conduct a survey on their own to measure corporate reputation. The Fortune magazine's survey is conducted annually among more than 8000 managers and financial analysts. It rates around 700 companies according to their innovativeness, people management, use of corporate assets, social responsibility, global competitiveness, quality of management, financial soundness, value as a long-term investment and product quality or service quality.

The use of surveys for measuring corporate reputation should be assessed critically because it does not cover all stakeholder groups of a company. It has been shown that the resulting reputation ratings reflect mostly the perception of the financial perspectives of a company [9], [10]. Thus, the meaningfulness of such ratings is limited. Furthermore, the low update frequency of the reputation index and the limitation to 700 companies reduces its usability further. Conducting an own survey is costly, time consuming and often covers stakeholders only partly (e.g., [3], [8]).

Nowadays, online media represent a very good source for reputation related comments by customers of companies. However, measuring corporate reputation from online media is a dynamic and challenging problem. The Internet in general extends the reach, speed and intensity of news [11]. There is a great number of online media outlets where people express their opinions about corporations and their products. Because of the volumes of textual data, manual processing is practically impossible. Furthermore, numerous factors that influence reputation need to be considered. However, an automatic retrieval approach using textual content from online media would be an efficient and holistic way to measure corporate reputation.

We propose to combine an information retrieval approach with sentiment analysis methods for automatically analyzing corporate reputation in online media. We contribute an ontology for identifying corporations in the first place. For

analyzing corporate reputation in text, we contribute a unique dataset of human annotated reputation texts. We use the dataset for corporation-specific reputational sentiment analysis using a machine learning classifier. Our work helps corporations to efficiently measure reputation, which is an important factor for the performance of a corporation.

The remainder of this paper is organized as follows. In section 2, corporate reputation is defined and approaches for measuring corporate reputation are presented. Section 3 specifies the problem. In section 4, the proposed reputation measurement approach is described. In section 5, we present our dataset of annotated reputation texts and evaluate our reputational sentiment classifier. Section 6 concludes.

## II. RELATED WORK

### A. Defining Corporate Reputation

Corporate reputation has been a popular topic in different streams of research, leading to a large amount of definitions of corporate reputation (e.g. [12-16]). According to [13,14] we do not use the terms corporate reputation, image and identity interchangeably. Based on [12-16] we define corporate reputation as the collective perception and judgment of the sentiment (i.e., feeling, opinion) about a corporation and its products or services by its stakeholders. Corporate reputation can be positive or negative [17]. Corporate reputation arises from the ability of a corporation to uphold social and institutional norms and values and to satisfy the needs and desires of its stakeholders. Corporate reputation forms through the appealing “character” [15] of a corporation and in the comparison with other entities.

### B. Approaches for Measuring Corporate Reputation

Most of the empirical reputation research uses the Fortune’s magazine “Most Admired Companies” (FMAC) index for measuring reputation [18]. It is based on a survey of senior executives and directors conducted annually. Companies with revenue of at least 10 billion \$ and at least the 15-th biggest revenue in their industry are ranked according to 9 “attributes of reputation”. The use of Fortune’s reputation data is rightfully criticized because it was shown to mostly reflect only the financial performance of a corporation [9], [10]. Surveying only senior executives and directors neglects all other stakeholder groups. The FMAC index also suffers of industry effects because the surveyed managers are explicitly asked to rate the corporations in comparison only to the other corporations in a particular industry [19]. The limited availability and frequency of reputation data (i.e., the reputation index refers to only the largest corporations) further limits the use of Fortune’s index for operative decision making.

“Britain’s most admired companies” (BMAC) of Management Today offers another publicly available reputation index. It is structurally very similar to the FMAC [20]. Similarly to FMAC, managers rate companies according to nine [20]. The critique to FMAC largely applies to BMAC as well because of the similarities between the two surveys.

“Reputation Quotient” (RQ) is a reputation ranking of the 60 “most visible” companies in the U.S. [21]. The companies are rated on 20 attributes distributed over six components of corporate reputation [21]. 22480 randomly selected respondents’ rated one or two companies. Each company is rated by at least 279 people. RQ is theoretically more founded than FMAC/BMAC but its commercial orientation complicates a closer examination. The fact that only the 60 most popular companies *at the time* are rated limits the usability both for research and practice because of resulting gaps in the time series and the small amount of observations.

Reputation can be also measured by conducting an own survey. This technique was employed by [3], [8], [22-24]. Modifications of the classical written (online) surveys like Verbal Protocol Analysis (taping, coding and analyzing the answers of respondents) [23] and the use of personification metaphor (rating a corporation on a five-point scale in regard to 42 items that load onto five orthogonal character factors) [25] have also been proposed.

### C. Research Gap

The reviewed studies on corporate reputation measurement have one major flaw: they do not cover all relevant stakeholder groups. This fact draws attention to the difficulty of conducting a representative survey of corporate reputation: it is very costly and time consuming. Conducting such a survey on a regular basis and for many corporations is practically impossible for smaller corporations.

We propose a different approach to measure corporate reputation in an automatic, efficient, and more holistic way by retrieving corporate reputation-related textual content from online media and using a sentiment analysis approach.

## III. PROBLEM SPECIFICATION

A document from online media may express reputational sentiments on multiple corporations [26]. Sentiments referring to multiple corporations can have different sentiment polarities. The problem is to classify the reputation sentiment polarity contained in a document with respect to each sentiment object separately. By classifying reputation sentiment, all factors influencing reputation should be considered.

## IV. APPROACH

This section describes our approach for extracting corporate reputation from online media texts. A machine learning based classifier is used for reputational sentiment classification. This approach does not require costly and time-consuming optimization of a knowledge base [29]. It is computationally efficient due to the linear classifier [30].

First, each document was pre-processed with natural language processing techniques, similarly to [27]. The preprocessing includes tokenization, sentence splitting, part of speech (POS)-tagging, and morphological analysis for lemmatization. Following [27, 28] the pre-processing was

implemented by GATE’s information extraction system [31].

The pre-processing includes ontology-based entity recognition. For this purpose, an ontology was developed based on [27]. The ontology contains all corporations from the Dow Jones Industrial Average, S&P 500 and S&P 600 indices, and various European and US banks that are also present in our reputation text dataset. For each corporation, hand-curated labels were defined for textual identification.

Second, we extracted relevant text segments that refer to a certain corporation, which were identified by the ontology. Following [32], the relevant text segment is defined as 25 words either side of the mention of a corporation. Then, all text segments referring to the same corporation within one article are concatenated.

Third, a linear kernel soft-margin Support Vector Machine (SVM) is applied successively on each of these text sections, as it has been shown to perform text classification tasks on state-of-the-art level, when given limited training data [30], [33]. We used the default hyperparameter configuration of SVM. The hyperparameter of the SVM for the costs associated with allowing training errors was set to 1. The features used by SVM are frequency counts of unigrams in a document [34]. The feature space of SVM contains only tokens of type “word” normalized by root (i.e., lemmatized words). Feature selection has not been used [35]. The result is a corporation-specific reputational sentiment (positive / negative) on document level.

## V. EVALUATION

The evaluation compares the classifier’s results to the gold standard, provided by a dataset of reputation-related texts, which have been annotated by humans for reputational sentiment.

The dataset consists of 688 text documents from online media related to corporations’ reputation. The documents were annotated by reputation experts from the banking sector, considering all factors that can influence reputation. Each document was annotated with a fuzzy sentiment label, i.e. each document was annotated with a specific degree of membership to the classes of positive and negative sentiment. The positive and negative membership degrees have five values each with an ascending degree of membership. In this work, binary annotations were derived from fuzzy sentiment labels by the following rule: documents with a higher positive than negative degree of membership is part of the positive class and all other document are part of the negative class. The positive class contains 40% of the documents of our dataset and the negative class 60%.

The dataset was annotated in three rounds: The first round consisted of 269 documents and was annotated by three experts. The dataset was randomly divided among the annotators so that each document was annotated by at least one annotator. In the second round, 394 documents were annotated by four annotators. Again, each document was classified by at least one randomly chosen annotator. In a third round, one annotator annotated 25 documents.

TABLE I.  
CLASSIFIER PERFORMANCE

	Precision	Recall	F-Measure	Accuracy
Positive	87.8%	70.8%	78.4%	84.4%
Negative	84.2%	91.1%	87.5%	84.4%
Micro Avg.	85.4%	83.0%	84.2%	

To evaluate the agreement among annotators for the reputational sentiment annotations, Fleiss’ Kapa inter-rater agreement for nominal scaled values with more than two raters was used [36]. In the first round, 27 documents have been annotated by all three annotators and considering only the positive and negative class, Fleiss’ Kappa of these annotations is 0.78. In the second round, all annotators annotated each of 49 documents. The Fleiss’ kappa from these 49 documents’ annotations is 0.66. We consider the level of agreement fairly well, thus the corpus can be used for evaluation of our classification approach.

Following [38, 37], stratified ten-fold cross validation was used. After classifying every document with the classifier on a test subset, we calculated the standard information retrieval metrics and micro averaged them [39].

Table 1 shows the evaluation results. Our approach could not recognize corporations or sentiment in 19 documents, which were not included in the evaluation. Our accuracy of 84.4% is comparable to results from state of the art sentiment classification research [35], [40].

## VI. DISCUSSION

The contribution of this work is an information retrieval approach for efficiently and comprehensively analyzing corporate reputation automatically from online media texts. Our approach builds upon an ontology for identifying all text parts relating to the same corporation. We contribute a unique dataset of labelled corporation reputation texts (see <https://wi2.uni-hohenheim.de/analytics>) and use it for corporation-specific reputational sentiment analysis by a machine learning method. The evaluation of our approach shows an overall accuracy of 84.4%.

A limitation of this work is that the neutral sentiment orientation is omitted, because [32] have found sentiment classification performance to be substantially higher when omitting the neutral class. We deliberately did not use deep learning techniques because the size of our dataset is too small. That is, the size of our dataset is a limitation. However, human annotation is costly and the size of our dataset is not much smaller than related work (e.g., [32]).

From a managerial perspective, our work helps to efficiently measure corporate reputation in on online world where news and opinions travel fast. Thus, managers can make better decisions by constantly monitoring reputation.

Future work points to comparing our measure of corporate reputation for online media with existing survey-based measures to gain insights about measurement validity. Furthermore, our measure of corporate reputation should be empirically validated by its ability to sense impacts on the financial and economic prospects of a corporation.

## REFERENCES

- [1] A. Pharoah, "Corporate Reputation: The Boardroom Challenge," *Corp. Gov.*, vol. 3, no. 4, pp. 46–51, 2003. <http://dx.doi.org/10.1108/14720700310497113>
- [2] H. Shefrin and M. Statman, "Making sense of beta, size and book-to-market," *J. Portfolio Manage.*, vol. 21, no. 2, pp. 26–34, 1995. <http://dx.doi.org/10.3905/jpm.1995.409506>
- [3] H. Shefrin, "Do investors expect higher returns from safer stocks than from riskier stocks?," *J. Psychol. Financ. Market.*, vol. 2, no. 4, pp. 37–41, 2001. [http://dx.doi.org/10.1207/S15327760JPFM0204\\_1](http://dx.doi.org/10.1207/S15327760JPFM0204_1)
- [4] D. MacGregor, P. Slovic, D. Dreman, and M. Berry, "Imagery, affect, and financial judgment," *J. Psychol. Financ. Market.*, vol. 1, no. 2, pp. 104–110, 2000. [http://dx.doi.org/10.1207/S15327760JPFM0102\\_2](http://dx.doi.org/10.1207/S15327760JPFM0102_2)
- [5] S. Hammond and J. Slocum, "The impact of prior firm financial performance on subsequent corporate reputation," *J. Bus. Ethics.*, vol. 15, no. 2, pp. 159–165, 1996. <https://doi.org/10.1007/BF00705584>
- [6] M. Sobol and G. Farrelly, "Corporate reputation: A function of relative size or financial performance," *Rev. Bus. Econ. Res.*, vol. 24, no. 1, pp. 45–59, 1988.
- [7] P. Roberts and G. Dowling, "Corporate reputation and sustained superior financial performance," *Strateg. Manage. J.*, vol. 23, no. 12, pp. 1077–1093, 2002. <http://dx.doi.org/10.1002/smj.274>
- [8] J. Bloemer, K. De Ruyter, and P. Peeters, "Investigating drivers of bank loyalty: the complex relationship between image, service quality and satisfaction," *Int. J. Bank. Market.*, vol. 16, no. 7, pp. 276–286, 1998. <https://doi.org/10.1108/02652329810245984>
- [9] G. E. Fryxell and J. Wang, "The Fortune Corporate 'Reputation' Index: Reputation for What?," *J. Manage.*, vol. 20, no. 1, pp. 1–14, 1994. <https://doi.org/10.1177/014920639402000101>
- [10] S. Brown, B. Perry, "Removing the Financial Performance Halo from Fortune's 'Most Admired' Companies," *Acad. Manage. J.*, vol. 37, no. 5, pp. 1347–1359, 1994. <https://doi.org/10.5465/256676>
- [11] V. Kubitschek, "Business discontinuity – a risk too far," *Balance Sheet*, vol. 9, no. 3, pp. 33–38, 2001. <http://doi.org/10.1108/09657960110696032>
- [12] C. J. Fombrun and C. B. M. van Riel, "The Reputational Landscape," *Corporate Reputation Review*, vol. 1, no. 1, pp. 5–13, 1997. <https://doi.org/10.1057/palgrave.crr.1540008>
- [13] M. L. Barnett, J. M. Jermier, and B. Lafferty, "Corporate Reputation: The Definitional Landscape," *Corporate Reputation Review*, vol. 9, no. 1, pp. 26–38, 2006. <http://doi.org/10.1057/palgrave.crr.1550012>
- [14] T. J. Brown, P. A. Dacin, M. G. Pratt, and D. . Whetten, "Identity, Intended Image, Construed Image, and Reputation: An Interdisciplinary Framework and Suggested Terminology," *J. Acad. Market. Sci.*, vol. 34, no. 2, pp. 99–106, 2006. <http://doi.org/10.1177/0092070305284969>
- [15] E. G. Love and M. Kraatz, "Character, Conformity, or the Bottom Line? How and Why Downsizing Affected Corporate Reputation," *Acad. Manage. J.*, vol. 52, no. 2, pp. 314–335, 2009. <http://doi.org/10.5465/AMJ.2009.37308247>
- [16] D. Lange, P. M. Lee, and Y. Dai, "Organizational Reputation: A Review," *J. Manage.*, vol. 37, no. 1, pp. 153–184, 2010. <http://doi.org/10.1177/0149206310390963>
- [17] P. Rhee, M. Haunschild, "The liability of good reputation: A study of product recalls in the US automobile industry," *Organization Science*, vol. 17, no. 1, pp. 101–117, 2006. <https://doi.org/10.1287/orsc.1050.0175>
- [18] D. Basdeo, K. Smith, C. M. Grimm, V. P. Rindova, and P. J. Derfus, "The impact of market actions on firm reputation," *Strateg. Manage. J.*, vol. 27, no. 12, pp. 1205–1219, 2006. <http://doi.org/10.1002/smj.556>
- [19] C. Fombrun and M. Shanley, "What's in a Name? Reputation Building and Corporate Strategy," *Acad. Manage. J.*, vol. 33, no. 2, pp. 233–258, 1990. <http://doi.org/10.2307/256324>
- [20] S. J. Brammer and S. Pavelin, "Corporate Reputation and Social Performance: The Importance of Fit," *Journal of Management Studies*, vol. 43, no. 3, pp. 435–455, 2006. <https://doi.org/10.1111/j.1467-6486.2006.00597.x>
- [21] C. Fombrun, "Corporate Reputation—its Measurement and Management," *Thesis*, vol. 18, no. 4, pp. 23–26, 2001.
- [22] D. Turban, D., Greening, "Corporate Social Performance and Organizational Attractiveness to prospective employees," *Acad. Manage. J.*, vol. 40, no. 3, pp. 658–672, 1997. <https://doi.org/10.5465/257057>
- [23] D. Cable and M. Graham, "The determinants of job seekers' reputation perceptions," *J. Organ. Behav.*, vol. 21, no. 8, pp. 929–947, 2000. [https://doi.org/10.1002/1099-1379\(200012\)21:8<929::AID-JOB63>3.0.CO;2-O](https://doi.org/10.1002/1099-1379(200012)21:8<929::AID-JOB63>3.0.CO;2-O)
- [24] V. Rindova and I. Williamson, "Being good or being known: An empirical examination of the dimensions, antecedents, and consequences of organizational reputation," *Acad. Manage. J.*, vol. 48, no. 6, pp. 1033–1049, 2005. <https://doi.org/10.5465/amj.2005.19573108>
- [25] G. Davies, R. Chun, and R. da Silva, "The personification metaphor as a measurement approach for corporate reputation," *Corporate Reputation Review*, vol. 4, no. 2, pp. 113–127, 2001. <https://doi.org/10.1057/palgrave.crr.1540137>
- [26] B. Liu and Zhang, "A survey of opinion mining and sentiment analysis," in *Mining Text Data*, 2012, pp. 415–463. [https://doi.org/10.1007/978-1-4614-3223-4\\_13](https://doi.org/10.1007/978-1-4614-3223-4_13)
- [27] A. Klein, O. Altuntas, T. Haeusser, and W. Kessler, "Extracting Investor Sentiment from Weblog Texts: A Knowledge-based Approach," in 13th Conference on Commerce and Enterprise Computing IEEE, 2011, pp. 1–9. <https://doi.org/10.1109/CEC.2011.10>
- [28] A. Klein, O. Altuntas, M. Riekert, and V. Dinev, "A Combined Approach for Extracting Financial Instrument-Specific Investor Sentiment from Weblogs," in 11th International Conference on Wirtschaftsinformatik, 2013, pp. 691–705.
- [29] F. Sebastiani, "Machine learning in automated text categorization," *ACM Computing Surveys*, vol. 34, no. 1, pp. 1–47, Mar. 2002. <https://doi.org/10.1145/505282.505283>
- [30] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," in 10th European Conference on Machine Learning, 1998, vol. 1398, no. 2, pp. 137–142. <https://doi.org/10.1007/BFb0026683>
- [31] D. Maynard et al., "Architectural elements of language engineering robustness," *Natural Language Engineering*, vol. 8, pp. 257–274, 2002. <https://doi.org/10.1017/S1351324902002930>
- [32] N. O'Hare et al., "Topic-Dependent Sentiment Analysis of Financial Blogs," in International CIKM Workshop on Topic-Sentiment Analysis for Mass Opinion Measurement, 2009, pp. 9–16. <https://doi.org/10.1145/1651461.1651464>
- [33] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," in Conference on Empirical Methods in Natural Language Processing, 2002, pp. 79–86. <https://doi.org/10.3115/1118693.1118704>
- [34] M. Riekert, J. Leukel, and A. Klein, "Online Media Sentiment: Understanding Machine Learning-Based Classifiers," *Proceedings of the 24th European Conference on Information Systems (ECIS)*, 2016.
- [35] H. Tang, S. Tan, and X. Cheng, "A survey on sentiment detection of reviews," *Expert Systems with Applications*, vol. 36, no. 7, pp. 10760–10773, Sep. 2009. <https://doi.org/10.1016/j.eswa.2009.02.063>
- [36] J. L. Fleiss, "Measuring nominal scale agreement among many raters," *Psychological bulletin*, vol. 76, no. 5, 1971. <http://doi.org/10.1037/h0031619>
- [37] B. Efron, "Estimating the error rate of a prediction rule: improvement on cross-validation," *Journal of the American Statistical Association*, vol. 78, no. 382, pp. 316–331, 1983. <https://doi.org/10.2307/2288636>
- [38] R. Kohavi, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," in *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, 1995, pp. 1137–1143.
- [39] Y. Yang, "An evaluation of statistical approaches to text categorization," *Information retrieval*, vol. 1, no. 1–2, pp. 69–90, 1999. <https://doi.org/10.1023/A:1009982220290>
- [40] R. Moraes, J. F. Valiati, and W. P. Gavião Neto, "Document-level sentiment classification: An empirical comparison between SVM and ANN," *Expert Systems with Applications*, vol. 40, no. 2, pp. 621–633, Feb. 2013. <https://doi.org/10.1016/j.eswa.2012.07.059>

# A Specialized Evolutionary Approach to the bi-objective Travelling Thief Problem

Maciej Laszczyk

Wrocław University of Science and Technology  
Faculty of Computer Science and Management  
ul. Ignacego Łukasiewicza 5, 50-371 Wrocław, Poland  
maciej.laszczyk@pwr.edu.pl

Paweł B. Myszkowski

Wrocław University of Science and Technology  
Faculty of Computer Science and Management  
ul. Ignacego Łukasiewicza 5, 50-371 Wrocław, Poland  
pawel.myszkowski@pwr.edu.pl

**Abstract**—In the recent years, it has been shown that real world-problems are often comprised of two, interdependent subproblems. Often, solving them independently does not lead to the solution to the entire problem. In this article, a Travelling Thief Problem is considered, which combines a Travelling Salesman Problem with a Knapsack Problem. A Non-Dominated Sorting Genetic Algorithm II (NSGA-II) is investigated, along with its recent modification - a Non-Dominated Tournament Genetic Algorithm (NTGA). Each method is investigated in two configurations. One, with generic representation, and genetic operators. The other, specialized to the given problem, to show how the specialization of genetic operators leads to better results. The impact of the modifications introduced by NTGA is verified. A set of Quality Measures is used to verify the convergence, and diversity of the resulting PF approximations, and efficiency of the method. A set of experiments is carried out. It is shown that both methods work almost the same when generic representation is used. However, NTGA outperforms classical NSGA-II in the specialized results.

## I. INTRODUCTION

**N**P-HARD optimization problems occur in many real-world scenarios. Be it a lot-sizing problem in economics [1], a transportation problem [2], or a scheduling problem [3]. These problems are ubiquitous and very practical, which makes their solving an important task. In practice, a problem often has multiple objectives. In scheduling problems, both time and cost of the schedule can be considered. In finance, it is desired to maximize the profits, but also to minimize the potential risks. Hence, multi-objective approaches aim to find a set of equally-good solutions, called a Pareto Front (PF).

Recently, authors of paper [4] have pointed out, that real-world problems comprise of multiple subproblems. They contain many dependencies and interwovenness. For that reason, it is not sufficient to find the solution to each of the subproblems. Objectives are interconnected in a way, that the improvement to one objective can lead to the worse value of another objective. Hence, combinations of such solutions do not guarantee the optimal solution to the entire problem. Authors of [5] proposed a Travelling Thief Problem (TTP), which has the features of a real-world problem. In this article, it is used in carried out experiments. It comprises of two constrained problems - a Travelling Salesman Problem (TSP) and a Knapsack Problem (KP). They are interconnected in a way, that makes solving them separately ineffective.

Due to its interdependence and multi-objective nature, evolutionary approaches show great potential in solving the TTP [6], [7], [8]. In [4] it has been shown that, in case of TTP, classical Non-Dominated Sorting Genetic Algorithm (NSGA-II) [9] with specialized operators outperforms other methods. Authors of [10] have introduced a Non-Dominated Tournament Genetic Algorithm. It is based on a NSGA-II, but contains multiple modifications. The authors carry out the research on a scheduling problem to show that these modifications lead to increased effectiveness of the method. This article attempts to verify the effectiveness of modifications introduced by NTGA with the combination of operators specialized for TTP.

A set of quality measures (QMs) proposed in [11] is used to evaluate the multi-objective results. Convergence and diversity of the resulting PF approximations are measured, along with the efficiency of the method.

The rest of the article is structured as follows. Section II presents existing work related to the subject. The TTP is described in section III. Section IV presents both used methods, as well as generic, and specialized representation, and operators. Results of all experiments along with the visualizations are presented in section V. Moreover, a theoretical analysis of the results is given. Lastly, section VI presents the conclusion and outlines the future work.

## II. RELATED WORK

The TTP was first introduced in [5]. The authors pointed out the shortcomings of benchmark problems used in the literature. The important features of a real-world problems were identified, namely existence of the subproblems and their interwovenness. Eventually, a single- and multi-objective versions of TTP were proposed.

Authors of [12] introduced a benchmark dataset for TTP. It contains 9720 instances. Each instance contains a TSP and KP elements. Additionally, the items are assigned to the cities to create the TTP instance. There are three different weight-value correlations present in the dataset. Moreover, instances contain up to 10 items per city.

Many researchers tackled the single-objective version of TTP. Authors of [13] have proposed three exact algorithms based on dynamic programming, branch and bound, and

constraint programming. Moreover, they compared them to the state-of-the-art solvers. In [8] swarm intelligence was used. Additionally, the authors investigated a TTP-specific local search algorithm. A Genetic Algorithm was used in [14]. Authors solve the overall problem instead of solving the subproblems separately. Moreover, the initial population is generated using a TSP specific heuristic. Authors of [18] used a hyperheuristic approach to select the best combination of known heuristics to solve the problem.

A multi-objective approach to TTP is considered less common in literature. However, authors of [15] used a combination of evolutionary computation and dynamic programming for the bi-objective TTP. Additionally, novel indicators were proposed, and the approach was compared to state-of-the-art methods. In [4], an NSGA-II with specialized representation and genetic operators was investigated. Various crossover and mutations method were investigated and the best configuration was identified. The results were compared to Greedy-based approaches. It was shown that NSGA-II outperforms other investigated methods.

NTGA was first proposed in [10]. The authors researched it on a bi-objective scheduling problem. The results were compared to classical NSGA-II and a decomposition based approach.

### III. PROBLEM

TTP is a constrained, combinatorial, NP-hard optimization problem. It comprises of two interwoven subproblems, namely TSP and KP. In the TSP part of the problem, there is a set of cities. Each city must be visited exactly once. In each of the cities there is a set of items, where each item has a weight and a value. Those items represent the KP part of the problem. While travelling a decision must be made which items to pick (if any).

TTP is a bi-objective problem. On one hand, the the goal is to find the quickest route between the cities. On the other, the total value of picked items must be maximized. However, each picked item decreases the speed of travel based on its weight. Hence, an improvement of the profit leads to an increase of the travelling time. TTP can be formally defined by equations 1 and 2.

$$\min f_{\tau}(\pi, z) \quad (1)$$

$$\max f_P(z) \quad (2)$$

where  $\pi$  is the permutation vector of all visited cities, and  $z$  is the picking plan.

The interaction between the subproblems is defined by equation 3. The total travelling time calculated as the sum of travelling times between each pair of consecutive cities plus the travelling time back to the first city. Each of those travelling times is influenced by all the picked items up to the given city.

$$f_{\tau}(\pi, z) = \sum_{i=1}^{n-1} \frac{d_{\pi_i, \pi_{i+1}}}{v(w(\pi_i))} + \frac{d_{\pi_n, \pi_1}}{v(w(\pi_n))} \quad (3)$$

$d_{\pi_i, \pi_{i+1}}$  is the distance between two consecutive cities from the permutation vector.  $n$  is the number of all cities.  $v(w(\pi_i))$  is the velocity in city  $\pi_i$ , considering the current weight  $w$  of picked items, and is defined by equation 4.

$$v(w) = v_{max} - \frac{W_c}{W}(v_{max} - v_{min}) \quad (4)$$

$W_c$  is the current weight, which is the sum of weights of all currently picked items.  $W$  is the capacity of the knapsack.  $v_{max}$  and  $v_{min}$  define the maximum and minimum allowed speed respectively.

The second objective, total profit, is the sum of values of all picked items. It is described by equation 5.

$$f_P(z) = \sum_{j=1}^m z_j z_j^{profit} \quad (5)$$

$m$  is the number of all items,  $z_j$  defines whether  $j$ 'th item has been picked and is equal to either 0 or 1.  $z_j^{profit}$  is the profit of  $j$ 'th item.

Additional constraints must be satisfied for the TTP solution to be *feasible*. The route must contain all cities, and each city must be visited exactly once. The sum of weights of all picked items cannot be greater than the capacity of the knapsack.

### IV. APPROACH

All approaches researched in the article are described in this section. First, definitions of important terms are given, namely dominance relation and a Pareto Front (PF). Next subsection describes used representation of an individual and is divided into two parts. First, a generic representation is described, and then one specialized for TTP. Similarly, the description of genetic operators starts with the operators used with generic representation. Later, the description of operators specialized for TTP is provided. Eventually, descriptions of NSGA-II and NTGA close out the section.

#### A. Definitions of Terms

This subsection contains a description of important terms. They are relevant for all used approaches.

1) *Dominance Relation*: One of the challenges of multi-objective optimization is the comparison of two solutions. Each solution is described by more than one objective, hence a numerical comparison is not sufficient. A Dominance Relation is defined for that purpose. A solution dominates another, if it has the value of at least one objective better and value of no objective worse, that that solution.

2) *Pareto Front*: A true PF contains all globally non-dominated solutions. However, in practice, a true PF is often not known. Hence, in this article, PF refers to the approximation found by the method, which contains all found non-dominated solutions.

### B. Representation

An individual used in evolutionary algorithms consists of a vector of numbers, a genotype. It represents the solution to the given problem. In this article, two representations are used. A generic one, presented in [10], and specialized for TTP, presented in [4].

1) *Generic*: The genome comprises two parts. The first one represents the solution to TSP. It assigns each city a priority. Then, the travelling plan is built by ordering the cities by that priority. The second part of the genome defines the solution to KP. It contains the number of genes equal to the number of items. Each item is assigned either 0 or 1, which defines whether the item should be picked or not.

For example a genome [1, 3, 3, 2, 0, 0, 1, 1] for a problem with 4 cities and 4 items means that the first city is visited first, then the last one, then second, and then third. Additionally, only the last 2 items are picked.

2) *Specialized*: Specialized representation defines the travelling plan with a permutation vector of all cities. The second part of genome is the same as in the case of generic representation.

For example a genome [1, 2, 4, 3, 1, 0, 0, 1] for a problem with 4 cities and 4 items means that the the first city is visited first, then the second one, then the last one, and finally the third one. The first and the last items are picked.

### C. Genetic Operators

This section describes the genetic operators used in the methods, for both representations.

Initial population generation is common for all methods and does not depend on the representation. A random initialization is used.

Selection operator is similarly independent of the representation, but is different for each method and is described in the appropriate method section (IV-D1 and IV-E1).

1) *Generic*: In case of generic representation, a standard single crossover and mutation operator is used for both parts of the genome. In the case of NSGA-II a single-point crossover is used. First, a random cut-point is selected within the genome. The first child is created by copying the genes on the left of that point from the first parent and copying the rest from the second parent. The second child is created similarly, by first copying the genes from the second parent and then from the first one. NTGA utilizes a single-point crossover.

Both methods use the same random mutation operator. First, a random gene is selected. Next, its value is randomly changed to a different, valid domain value.

2) *Specialized*: In the case of specialized representation, to include a problem domain knowledge a different crossover and mutation operator is used for each part of the genome. The Edge Operator (introduced in [16]) is used as the crossover for the part responsible for TSP. It aims to introduce as few as possible additional paths. It does so, by reusing existing edges when generating the children. First, a list is generated, which contains the neighbour cities from both parents for each city. Then, the first city from the first parent is copied to the

child genome and it is removed from the neighbour list. Then, iteratively, neighbours of that city with the fewest neighbours are copied over and removed from the list consecutively. If there are no more neighbours for a given city, a random city is selected. The second child is created similarly, by starting the process with the first city from the second parent. For the KP part of the genome, a uniform crossover is used. Children are created by copying each gene from a random parent with equal probability.

A swap mutation is utilized for the TSP part of the genome. Two random cities are selected and their position in the genome is swapped. For the KP part, a *bitflip* mutation is used. A random item is selected and the value of its gene is flipped.

In generic representation, there may occur a situation that makes an individual, a not *feasible* one. For example, some cities may have the same priority (cities are visited in defined order). In specialized representation crossover/mutation assure the feasibility of TSP-part of genome. However, another situation may exist in both representations when items picked by individual exceed knapsack capacity – items with min profit/weight ratio are removed from the solution.

### D. Non-Dominated Sorting Genetic Algorithm II

This section contains the description of a Non-Dominated Sorting Genetic Algorithm II (NSGA-II). It starts with the description of a pseudocode. Then, the selection and crowding distance, which are unique for this method are described.

NSGA-II is an evolutionary method. It processes a population of individuals in an iterative manner. Each individual represents a single solution to the given problem. The algorithm runs for the predefined number of generations, where a generation is a process of creating an offspring population from the current population. Each generation utilizes genetic operators to select parents and generate children individuals. Eventually, all non-dominated individuals found during the computation constitute a PF approximation. NSGA-II is described in pseudocode 1.

A *PopulationSize* parameter is stored in the first line. Then, in the second line, an initial population of that size is generated. In the third line, the entire population is evaluated. Each individual gets assigned the values of all objectives. Line 4 uses a non-dominated sorting to sort the population based on the rank and crowding distance, which are described in sections IV-D2 and IV-D3 respectively. In line 5, the loop begins, which iterates over all generations. Then, line 6 begins the loop to effectively double the size of the population. First, 2 parents are selected in line 7, using a selection described in IV-D1. In line 8, the crossover is used to create 2 children individuals. They are then mutated in line 9. Finally, the children are evaluated in line 10, and added to the current population in line 11. After the size of population has been doubled, it is again sorted in line 13. The population is truncated to its original size in line 14. Only the better half

**Algorithm 1** Pseudocode of NSGA-II [9]

---

```

1:  $N \leftarrow PopulationSize$ 
2:  $P_{current} \leftarrow generateInitialPopulation(N)$ 
    $evaluate(P_{current})$ 
3:  $nonDominatedSorting(P_{current})$ 
   for  $i \leftarrow 0$  to  $generationLimit$  do
4:   while  $|P_{current}| < 2N$  do
      $parents \leftarrow select(P_{current})$ 
5:    $children \leftarrow crossover(parents)$ 
      $children \leftarrow mutate(children)$ 
6:    $evaluate(children)$ 
      $P_{current} \leftarrow P_{current} \cup children$ 
7:   end while
    $nonDominatedSorting(P_{current})$ 
8:    $truncate(P_{current}, N)$ 
9:   end for
10: return  $P_{current}$ 

```

---

of the population remains. After all the generations have been processed, the last population is returned in line 16.

1) *Selection*: Selection in NSGA-II starts with taking 2 random individuals from the population. Then, those 2 individuals are compared based on the rank and crowding distance, which are described in sections IV-D2 and IV-D3 respectively. The individual with the lower rank is selected. If both have the same rank, individual with the larger crowding distance is selected. Selection returns only one parent, so during the algorithm it is performed twice, to obtain 2 parents.

2) *Rank*: During the NSGA-II each individual is assigned a rank, based on its quality. The lower rank means the individual is better. It is computed in an iterative manner. First, all non-dominated individuals are assigned the rank equal to 1. Then, all individuals that remain-dominated, while not considering the individuals with the rank already set, have the rank set to 2. The process is repeated, until all individuals are assigned a rank. Intuitively, the process divides the population into multiple PF approximations. The rank describes to which of those approximations the individual belongs.

3) *Crowding Distance*: Crowding distance is calculated for each individual. First, the largest possible box is drawn around the individual, that contains only that individual from the population. The crowding distance is the volume of that box. The larger values mean that the individual lies in the poorly explored part of the space. At the same time, lower values mean that there are many individuals around given individual.

### E. Non-Dominated Tournament Genetic Algorithm

This section contains the description of a Non-Dominated Tournament Genetic Algorithm (NTGA). First a pseudocode is given. Then its selection and clone elimination methods are described.

NTGA is based on a classical NSGA-II method. It introduces 4 modifications that aim to improve the effectiveness of the method. First, it separates parent and child populations. Then, it utilizes a selection method with stronger selective

pressure. Finally, it introduces a clone elimination method and archive usage. NTGA is presented in pseudocode 2.

**Algorithm 2** Pseudocode of NTGA [10]

---

```

1:  $N \leftarrow PopulationSize$ 
2:  $archive \leftarrow \emptyset$ 
    $P_{current} \leftarrow generateInitialPopulation(N)$ 
3:  $evaluate(P_{current})$ 
    $updateArchive(P_{current})$ 
4: for  $i \leftarrow 0$  to  $generationLimit$  do
    $P_{next} \leftarrow \emptyset$ 
5:   while  $|P_{next}| < |P_{current}|$  do
      $parents \leftarrow select_{tour}(P_{current})$ 
6:    $children \leftarrow crossover(parents)$ 
      $children \leftarrow mutate(children)$ 
7:   while  $P_{next}$  contains  $children$  do
      $children \leftarrow mutate(children)$ 
8:   end while
    $evaluate(children)$ 
9:    $P_{next} \leftarrow P_{next} \cup children$ 
    $updateArchive(children)$ 
10:  end while
    $P_{current} \leftarrow P_{next}$ 
11: end for
12: return  $archive$ 

```

---

First line stores the *PopulationSize* parameter. An empty archive is initialized in line 2. It is designed to store all found non-dominated individuals. In line 3, an initial population of given size is created. It is then evaluated in line 4. In line 5, the archive is updated with all currently non-dominated individuals. The loop, in line 6, iterate over a predefined number of generations. In line 7, an empty population is initialized, which is going to store the next population. The loop, in line 8, runs until the size of next population is equal to the size of current population. In line 9, parents are selected with the selection method described in IV-E1. Then, the children are created with the crossover in line 10. They are mutated in line 11. The clone elimination method is described between lines 12 and 13. If the next population already contains generated children, they are mutated. After that, the children are evaluated in line 15. Finally, they are added to the next population in line 16. In line 17 the archive is updated. The children are added to it, if they are non-dominated. Then, all individuals, that the children dominate, are removed from the archive. When the next population has been fully generated, it replaces the current population. At the end, the archive of non-dominated individuals is returned. The archive contains the PF approximation.

1) *Selection*: NTGA uses a tournament selection. First, given number of individuals is randomly drawn from the population. Then, they are compared according to their rank (described in IV-D2). The individual with the lowest rank is selected. If there are multiple individuals that match, the first one is selected. It is worth noting that the crowding distance is not considered during the comparison.

2) *Clone Elimination*: Clone elimination aims to increase the diversity of the population. The clones are defined as individuals with the identical genome. Before adding a child individual to the population, a check is performed, to verify whether an identical individual already exists. If so, the child is mutated until it is no longer a clone. Only then, it is added to the next population.

## V. EXPERIMENTS AND RESULTS

The experiments carried out in this article aim to verify the effectiveness of NTGA on TTP. Moreover, it is verified whether the modifications of NTGA are effective in the context of specialized operators. To answer those question the experiments on 4 configurations are carried out. NSGA-II and NTGA are researched with both generic, and specialized representation, and operators. A set of selected QMs is used to verify the convergence and diversity of the resulting PF approximations.

First, selected data instances and quality measures are described. Then, the experimental procedure is presented and selected parameter values are provided. A full set of experiments is carried out on all 4 configurations and the results are presented. Finally, the last subsection contains the theoretical analysis.

### A. Data Instances

A benchmark dataset, first presented in [12], is used in this article. In literature, an *eil51* instances are often used [17] [18], and so they have been selected for the research. 12 instances have been selected with 51 cities and the number of items between 50 and 500. 3 types of correlation between item weight and profit can be identified within the set. A strong correlation, where increased profit also means increased value. No correlation, but all the items have similar weights. The last group has no correlation between the items.

### B. Quality Measures

A set of QMs presented in [11] is used to verify the effectiveness of the methods. Convergence, diversity of the PF approximation, and the efficiency of the method is verified. A Perfect Point and a Nadir Point are used as a reference in 2 of the measures.

1) *Perfect Point*: A Perfect Point contains the best values of all objectives. It does not have to an achievable solution. The value of travelling time is calculated as the length of minimum spanning tree of the tour. A brute force search algorithm is used to calculate the value of profit.

2) *Nadir Point*: A Nadir Point contains the worst values of all objectives from among the non-dominated solutions. It often has to be approximated. Additionally, to make the comparison fair even worse values can be selected [19]. The value of travelling time is calculated by taking the value of travelling time from a Perfect Point and doubling it. It is an upper bound of the TSP. The profit is set to 0 and represents a solution where no items are picked.

3) *Euclidean Distance*: Euclidean Distance (*ED*) is a measure of convergence. It shows how close the PF approximation is to the true PF. Since the true PF is not known, *ED* utilizes the Perfect Point. Value of *ED* is obtained by calculating the average distance between every point on the PF approximation and the Perfect Point. It can be formally defined by equation 6.

$$ED(PF) = \frac{\sum_{i=1}^{|PF|} d_i}{|PF|} \quad (6)$$

*PF* is the Pareto Front,  $d_i$  is the distance from the  $i$ 'th point to the Perfect Point.

4) *Hypervolume*: Hypervolume (*HV*) is a measure of diversity. It is a volume of a hypercube defined by the Nadir Point and the PF approximation. It measures the spread, but is also influenced by the convergence and uniformity of the PF approximation. *HV* can be formally defined by equation 7.

$$HV(PF) = \Lambda\left(\bigcup_{s \in PF} \{s' | s \prec s' \prec s^{nadir}\}\right) \quad (7)$$

*PF* is an approximation of PF.  $s$  is the point of approximated PF.  $s^{nadir}$  is a *Nadir Point*.  $\Lambda$  is a Lebesgue measure, which is the generalization of a volume.  $\prec$  is a domination relation.

5) *Pareto Front Size*: Pareto Front Size (*PFS*) measures the diversity in terms of the cardinality of the PF approximation. It is defined as the number of points on the PF approximation.

6) *Ratio of Non-Dominated Individuals*: Ratio of Non-Dominated Individuals (*RNI*) measure the efficiency of the method. It is defined as the number of points on the PF approximation divided by the number of all visited points.

7) *Spacing*: Spacing (*S*) measures the uniformity of the PF approximation. It ensures that the solutions are evenly distributed and identifies the clustering effect. To calculate it, first, the distances between all consecutive points on the PF approximation are calculated. The standard deviation of those distances is the *S* measure. It can be defined with equation 8.

$$S(PF) = \sqrt{\frac{1}{|PF|} \sum_{i=1}^{|PF|} (d_i - \bar{d})^2} \quad (8)$$

*PF* is the approximation of the PF.  $d_i$  is the distance from the  $i$ -th point the next consecutive point.

The intuition of QMs is: *ED* should be minimized and measures the closeness to the true PF. *HV* should be maximized and it is influenced by both spread of the PF approximation and its distance to the true PF. *PFS* is simply the cardinality of the approximation, while *RNI* measures efficiency, by calculating the ratio of points on the approximation to all explored points. Spacing (*S*) should be minimized and it measures how closely the approximation resembles the uniform distribution.

### C. Experimental Procedure

First, the parameters of all methods have been tuned separately. Then, both NSGA-II and NTGA have been ran with both representations, on every instance. Due to the stochastic nature of evolutionary algorithms, each experiment has been repeated 50 times and results have been averaged. Next, a set of selected QMs has been calculated for each PF approximation. Statistical significance has been verified. Eventually, visualizations for selected instances are presented and theoretical analysis is described.

### D. Parameters

The first step in parameter tuning was to define a set of configurations of different parameter values. Taguchi Orthogonal Arrays have been used for that purpose. Then, each configuration has been ran 10 times and the QMs have been calculated for the resulting PF approximations. Next, a Multi-Objective Grey Relational Grade is calculated and a Taguchi Method is used to identify the impact of the parameters on the results [20]. Finally, the best parameter configuration is selected. This process has been repeated for all 4 configurations. The experiments show that a 1000 generations should be sufficient, however the limit has been set to 2000 to make sure the results converge.

Table I contains the selected configurations of parameter values for each researched method.

### E. Experiments

Tables II and III contain the results of experiments for NSGA-II and NTGA respectively. Both tables show the results on generic representation. Tables IV and V contain the results for NSGA-II and NTGA with specialized representation.

1) *Results*: Comparison of the results for the generic representation shows no significant difference between the results of QMs. The largest difference can be observed for *HV* for the benefit of NSGA-II, but it is within a single standard deviation.

Specialized operators improve the results significantly. The largest difference can again be observed for *HV*. NSGA-II with specialized operators have achieved better values of *HV* for all researched instances. Specialization has improved ED for 6 out of 12 instances. Overall, smaller instances show more improvement in terms of ED. Interestingly, *PFS* has been decreased almost 6 times. However, large values for generic representation might suggest that the solutions are far from local optima. There is no statistical difference in values of *RNI*. Values of *S* have also deteriorated. The largest difference can be seen for instances *eil51\_n250\_uncorr\_01* and *eil51\_n500\_uncorr\_01*. The difference can be justified by much larger spread of the approximation, which can be confirmed by the larger values of *HV*.

NTGA with specialized operators improves the results even further. On average, values of ED have been improved by almost 40%. The largest difference can be observed for larger instances. Values for *eil51\_n50\_bounded-strongly-corr\_01* and *eil51\_n50\_uncorr-similar-weights\_01* are worse than in case of NSGA-II. Better values of *HV* have been achieved for all

12 instances, which suggest much better diversity of the PF approximations. Similarly, larger *PFS* has been achieved for all 12 instances. For *eil51\_n250\_uncorr-similar-weights\_01* the value has been almost tripled. NTGA has also proved to be more efficient. It can be observed by larger values of *RNI*. The only instance that has not been improved is *eil51\_n50\_uncorr-similar-weights\_01*. Achieved values of *S* measure are also lower, however they are still almost 3 times larger than the values achieved for configurations with generic representation. Results compared to specialized versions of NSGA-II and NTGA presented in this section have been statistically confirmed by Wilcoxon signed-rank ( $W_{0.05} = 78 > W_c = 13$ ) for all QM's. All difference are statistically significant.

2) *Visualizations*: This section contains visualization for two selected instances. Figure 1 presents instance *eil51\_n50\_bounded-strongly-corr\_01*. It presents the case, where specialized NTGA has achieved the worse result than specialized NSGA-II in terms of ED measure. ED depends on a Perfect Point, which lies closest to the middle of PF. NSGA-II has achieved a large spread, but its approximation has very few points on the edges. Hence, average distance to the Perfect Point is relatively low. PF approximation generated by NTGA is more evenly distributed, and so many points lie far from the Perfect Point. Hence, ED deteriorates.

Figure 2 presents the second selected instance. For *eil51\_n500\_uncorr-similar-weights\_01* specialized NTGA improved the results the most in comparison to other configurations. Interestingly, specialized NSGA-II has generated the solutions with better profit than specialized NTGA.

Moreover, visualization of all achieved PF approximations for instance *eil51\_n250\_bounded-strongly-corr\_01* is presented in Figure 3. A modified version of empirical attainment function (EAF) [21] is used to get the "averaged" Pareto Front approximations. For clarity, only specialized configurations are shown. It can be seen, that NTGA achieves better results on average. Additionally, the deviations in the results are also smaller. However, NSGA-II has generated points with very high profit, that have not been dominated by any of the runs of NTGA.

### F. Summary

Table VI contains the summary of all obtained results for all configurations. NTGA does not improve the results in case of generic representation. Values of all measures are very similar for both NSGA-II and NTGA.

Specialization has improved the convergence and diversity of the PF approximation. It can be observed by the improved values of *ED* and *HV*. Specialization has decreased the value of *PFS*. However, in case of generic representation, achieved solutions are far from optimal, so their larger number is less significant. Specialized representation also led to increased distances between the points of the PF approximation. It might have been caused by the larger achieved spread.

In case of specialized representation, NTGA has improved the results significantly, even in comparison to specialized

TABLE I  
SELECTED PARAMETER CONFIGURATIONS

	representation	populationSize	generationLimit	$P_{m_{tsp}}$	$P_{m_{kp}}$	$P_{x_{tsp}}$	$P_{x_{kp}}$	tournamentSize
NSGA-II	generic	200	2000	0.005	0.005	0.9	0.9	2
	specialized	100	2000	0.1	0.05	0.6	0.8	2
NTGA	generic	50	2000	0.005	0.005	0.9	0.9	6
	specialized	50	2000	0.1	0.05	0.6	0.8	6

TABLE II  
VALUES OF SELECTED QMS FOR NSGA-II WITHOUT SPECIALIZATION

Instance	ED		HV		PFS		RNI		S	
	avg	std	avg	std	avg	std	avg	std	avg	std
eil51_n50_bounded-strongly-corr_01	0.4255	0.0486	0.7865	0.0297	65.4	8.3	0.0002	0.0000	0.0075	0.0029
eil51_n50_uncorr-similar-weights_01	0.3452	0.0338	0.6275	0.0122	3.3	1.6	0.0000	0.0000	0.0148	0.0109
eil51_n50_uncorr_01	0.3093	0.0319	0.8121	0.0064	34.0	8.2	0.0001	0.0000	0.0100	0.0025
eil51_n150_bounded-strongly-corr_01	0.3124	0.0156	0.7644	0.0198	169.0	46.6	0.0004	0.0001	0.0029	0.0011
eil51_n150_uncorr-similar-weights_01	0.3106	0.0390	0.6896	0.0586	47.4	27.6	0.0001	0.0001	0.0051	0.0025
eil51_n150_uncorr_01	0.2351	0.0130	0.7975	0.0131	90.6	25.2	0.0002	0.0001	0.0036	0.0011
eil51_n250_bounded-strongly-corr_01	0.2924	0.0111	0.7297	0.0173	206.1	57.1	0.0005	0.0001	0.0019	0.0009
eil51_n250_uncorr-similar-weights_01	0.3129	0.0175	0.7162	0.0152	97.3	29.5	0.0002	0.0001	0.0035	0.0012
eil51_n250_uncorr_01	0.2291	0.0108	0.8029	0.0160	144.5	33.1	0.0004	0.0001	0.0017	0.0005
eil51_n500_bounded-strongly-corr_01	0.2647	0.0119	0.7157	0.0186	253.7	101.1	0.0006	0.0003	0.0009	0.0002
eil51_n500_uncorr-similar-weights_01	0.3208	0.0190	0.6967	0.0139	162.1	57.2	0.0004	0.0001	0.0022	0.0010
eil51_n500_uncorr_01	0.2588	0.0086	0.7578	0.0076	189.9	56.5	0.0005	0.0001	0.0011	0.0006
Average	0.3014	0.0217	0.7414	0.0190	121.9	37.7	0.0003	0.0001	0.0046	0.0021

TABLE III  
VALUES OF SELECTED QMS FOR NTGA WITHOUT SPECIALIZATION

Instance	ED		HV		PFS		RNI		S	
	avg	std	avg	std	avg	std	avg	std	avg	std
eil51_n50_bounded-strongly-corr_01	0.4216	0.0451	0.7591	0.0330	55.4	13.5	0.0001	0.0000	0.0084	0.0052
eil51_n50_uncorr-similar-weights_01	0.3307	0.0247	0.6439	0.0155	5.3	2.2	0.0000	0.0000	0.0140	0.0103
eil51_n50_uncorr_01	0.2951	0.0290	0.8175	0.0103	31.0	8.4	0.0001	0.0000	0.0109	0.0042
eil51_n150_bounded-strongly-corr_01	0.3119	0.0147	0.7441	0.0218	144.3	52.0	0.0004	0.0001	0.0038	0.0016
eil51_n150_uncorr-similar-weights_01	0.3123	0.0334	0.6932	0.0383	49.1	25.2	0.0001	0.0001	0.0049	0.0019
eil51_n150_uncorr_01	0.2376	0.0118	0.7893	0.0123	86.6	28.2	0.0002	0.0001	0.0029	0.0008
eil51_n250_bounded-strongly-corr_01	0.2963	0.0140	0.7118	0.0110	202.4	79.0	0.0005	0.0002	0.0022	0.0012
eil51_n250_uncorr-similar-weights_01	0.3057	0.0195	0.7052	0.0166	93.0	39.8	0.0002	0.0001	0.0030	0.0010
eil51_n250_uncorr_01	0.2343	0.0148	0.7885	0.0143	116.3	40.5	0.0003	0.0001	0.0021	0.0021
eil51_n500_bounded-strongly-corr_01	0.2726	0.0097	0.6965	0.0171	263.0	105.6	0.0007	0.0003	0.0011	0.0011
eil51_n500_uncorr-similar-weights_01	0.3397	0.0276	0.6669	0.0335	159.9	73.5	0.0004	0.0002	0.0021	0.0010
eil51_n500_uncorr_01	0.2608	0.0185	0.7481	0.0165	214.8	51.9	0.0005	0.0001	0.0009	0.0004
Average	0.3016	0.0219	0.7303	0.0200	118.4	43.3	0.0003	0.0001	0.0047	0.0026

NSGA-II. Both ED and HV values have been improved. Additionally, efficiency of the algorithm has been improved, which can be observed by the increased value of RNI. Value of S measure has been improved in comparison to specialized NSGA-II. However, it remains higher than in case of generic representation.

### G. Theoretical Analysis

In NTGA, parent individuals do not have to compete with children individuals. There is no possibility that an individual will survive for multiple generations. Hence, more unique points are explored by the method. In combination with increased selective pressure it also leads to increased convergence. Interestingly, introduction of clone prevention has not improved the diversity. Larger values of HV are caused by the larger distance from the Nadir Point and not by the larger spread of the approximation. More significant improvement can be observed for larger instances.

Modifications of NTGA lead to no significant improvement in case of generic representation. Non-specialized operators have a low probability of improving the result. In consequence, increased selective pressure has much less significance.

The crowding distance has been removed from NTGA. Instead, each new individual has to be compared with the existing individuals to verify whether it is a clone. In most cases the comparison is done only once. However, if the individual is a clone, it is mutated and the check is performed again. In an edge case the comparison must be done multiple times, which might negatively affect the performance.

## VI. CONCLUSIONS AND FUTURE WORK

In this paper NTGA has been investigated in the context of TTP. A bi-objective problem, which comprises of two subproblems. The subproblems are interconnected, which makes solving them independently ineffective. NTGA has been compared to classical NSGA-II. All experiments have been carried out

TABLE IV  
VALUES OF SELECTED QMs FOR NSGA-II WITH SPECIALIZED REPRESENTATION

Instance	ED		HV		PFS		RNI		S	
	avg	std	avg	std	avg	std	avg	std	avg	std
eil51_n50_bounded-strongly-corr_01	0.3239	0.0484	0.8492	0.0092	30.2	6.3	0.0002	0.0000	0.0301	0.0140
eil51_n50_uncorr-similar-weights_01	0.4049	0.0592	0.7106	0.0067	9.8	2.7	0.0000	0.0000	0.0639	0.0286
eil51_n50_uncorr_01	0.2094	0.0172	0.8444	0.0091	11.6	3.5	0.0001	0.0000	0.0188	0.0072
eil51_n150_bounded-strongly-corr_01	0.2315	0.0240	0.8094	0.0131	33.5	8.4	0.0002	0.0000	0.0233	0.0299
eil51_n150_uncorr-similar-weights_01	0.2283	0.0289	0.7877	0.0127	16.0	7.3	0.0001	0.0000	0.0312	0.0469
eil51_n150_uncorr_01	0.2016	0.0506	0.8202	0.0101	16.5	4.5	0.0001	0.0000	0.0379	0.0642
eil51_n250_bounded-strongly-corr_01	0.2349	0.0492	0.8060	0.0169	31.7	10.2	0.0002	0.0001	0.0466	0.0459
eil51_n250_uncorr-similar-weights_01	0.2890	0.1075	0.7956	0.0138	16.9	8.7	0.0001	0.0000	0.0829	0.0801
eil51_n250_uncorr_01	0.2756	0.0942	0.8269	0.0099	18.4	5.4	0.0001	0.0000	0.1072	0.0785
eil51_n500_bounded-strongly-corr_01	0.3336	0.0805	0.7776	0.0157	33.4	11.1	0.0002	0.0001	0.0800	0.0397
eil51_n500_uncorr-similar-weights_01	0.4108	0.0836	0.8049	0.0094	26.9	10.2	0.0001	0.0001	0.0894	0.0253
eil51_n500_uncorr_01	0.4596	0.1021	0.8136	0.0090	25.0	8.1	0.0001	0.0000	0.1372	0.0395
Average	0.3003	0.0621	0.8038	0.0113	22.5	7.2	0.0001	0.0000	0.0624	0.0417

TABLE V  
VALUES OF SELECTED QMs FOR NTGA WITH SPECIALIZED REPRESENTATION

Instance	ED		HV		PFS		RNI		S	
	avg	std	avg	std	avg	std	avg	std	avg	std
eil51_n50_bounded-strongly-corr_01	0.3881	0.0424	0.8752	0.0098	51.3	10.6	0.0005	0.0001	0.0211	0.0078
eil51_n50_uncorr-similar-weights_01	0.4234	0.0738	0.7365	0.0039	12.9	4.2	0.0001	0.0000	0.0533	0.0145
eil51_n50_uncorr_01	0.2205	0.0221	0.8786	0.0052	19.2	4.3	0.0002	0.0000	0.0167	0.0063
eil51_n150_bounded-strongly-corr_01	0.2107	0.0190	0.8513	0.0112	67.1	18.1	0.0007	0.0002	0.0058	0.0025
eil51_n150_uncorr-similar-weights_01	0.2199	0.0373	0.8286	0.0105	37.5	17.8	0.0004	0.0002	0.0129	0.0060
eil51_n150_uncorr_01	0.1544	0.0062	0.8553	0.0065	31.6	8.8	0.0003	0.0001	0.0045	0.0026
eil51_n250_bounded-strongly-corr_01	0.1777	0.0151	0.8447	0.0119	67.8	16.4	0.0007	0.0002	0.0053	0.0063
eil51_n250_uncorr-similar-weights_01	0.1821	0.0183	0.8471	0.0096	53.9	18.3	0.0005	0.0002	0.0073	0.0028
eil51_n250_uncorr_01	0.1478	0.0068	0.8639	0.0063	33.9	8.8	0.0003	0.0001	0.0031	0.0025
eil51_n500_bounded-strongly-corr_01	0.1645	0.0103	0.8265	0.0102	70.6	17.8	0.0007	0.0002	0.0041	0.0070
eil51_n500_uncorr-similar-weights_01	0.1659	0.0131	0.8459	0.0098	64.9	22.0	0.0006	0.0002	0.0051	0.0047
eil51_n500_uncorr_01	0.1553	0.0114	0.8442	0.0081	36.5	9.9	0.0004	0.0001	0.0084	0.0294
Average	0.2175	0.0230	0.8415	0.0086	45.6	13.1	0.0005	0.0001	0.0123	0.0077

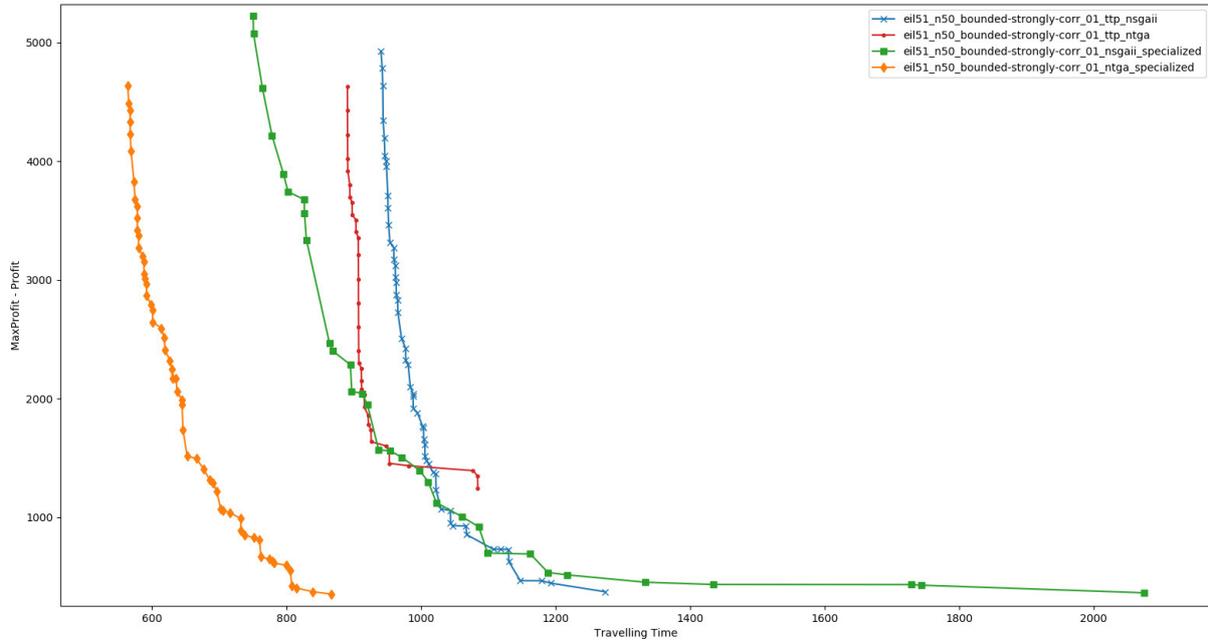


Fig. 1. Comparison of selected approx. Pareto Fronts for data instance eil51\_n50\_bounded-strongly-corr\_01

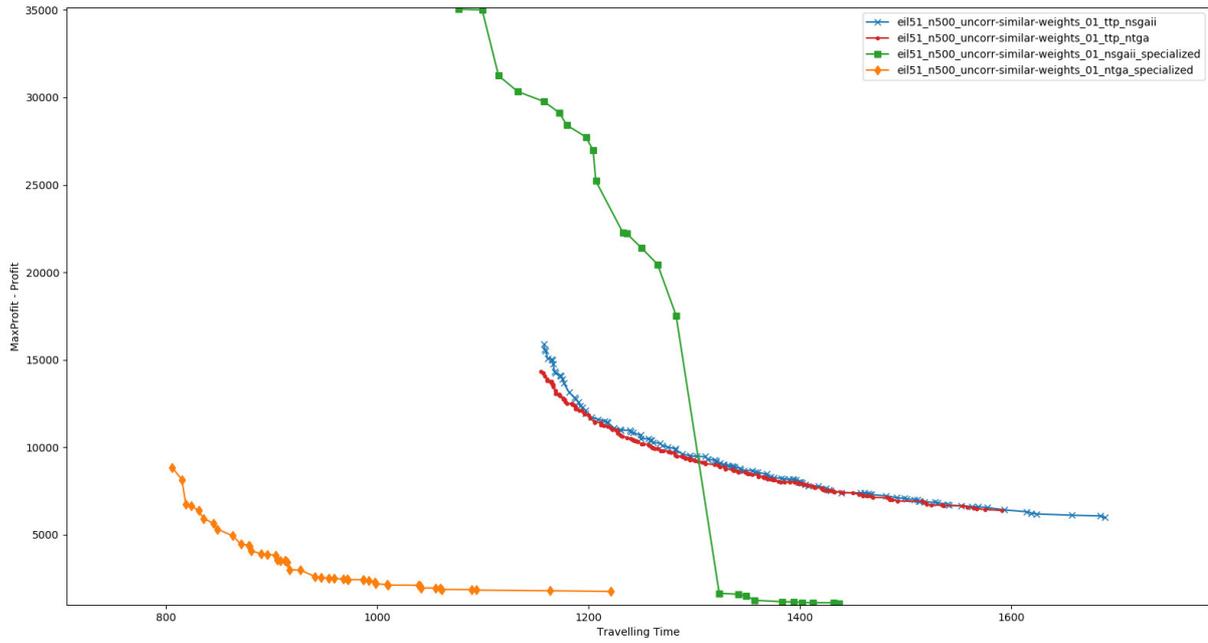


Fig. 2. Comparison of selected approx. Pareto Fronts for data instance eil51\_n500\_uncorr-similar-weights\_01

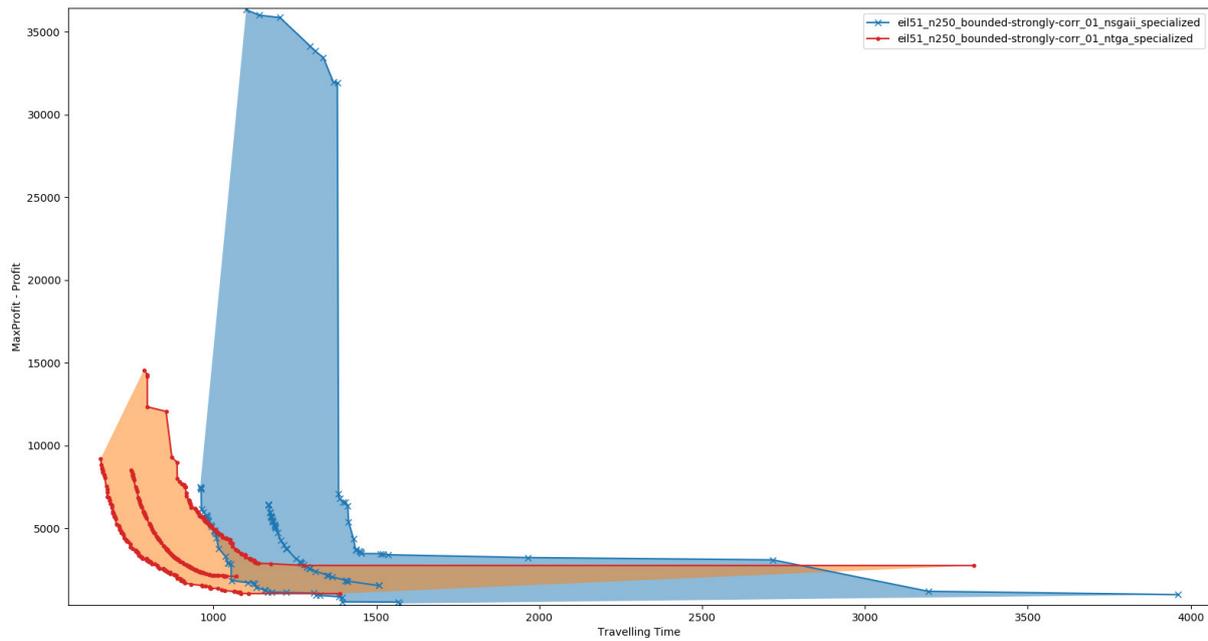


Fig. 3. Comparison of EAF average approx. Pareto Fronts for data instance eil51\_n250\_bounded-strongly-corr\_01

TABLE VI  
SUMMARY OF ALL RESULTS

		ED		HV		PFS		RNI		S	
		avg	std	avg	std	avg	std	avg	std	avg	std
NSGA-II	generic	0.3014	0.0217	0.7414	0.0190	121.9	37.7	0.0003	0.0001	0.0046	0.0021
	specialized	0.2991	0.0622	0.8029	0.0124	22.3	6.9	0.0002	0.0001	0.0659	0.0431
NTGA	generic	0.3016	0.0219	0.7303	0.0200	118.4	43.3	0.0003	0.0001	0.0047	0.0026
	specialized	0.2158	0.0208	0.8419	0.0084	45.4	12.2	0.0005	0.0001	0.0120	0.0079

in configurations with generic, and specialized representation. It has been shown, that for larger instances specialized NTGA achieves better results than specialized NSGA-II.

Increased selective pressure of NTGA led to improved results. However, more research could be done regarding selection, that would also promote diversity of the PF approximation. An introduction of heuristics, that would further improve the solutions for the subproblems might be worth investigating. Additionally, a hyperheuristic that would combine the benefits of multiple evolutionary methods could prove beneficial to the results. Moreover, many-objective problems are fairly uncommon. An interesting avenue of future work would be to use a benchmark problem with the real-world characteristics, with a larger number of objectives.

#### REFERENCES

- [1] Akbalik, Ayse, et al. "NP-hard and polynomial cases for the single-item lot sizing problem with batch ordering under capacity reservation contract." *European Journal of Operational Research* 257.2 (2017): 483-493.
- [2] Sanei, Masoud, et al. "Step fixed-charge solid transportation problem: a Lagrangian relaxation heuristic approach." *Computational and Applied Mathematics* 36.3 (2017): 1217-1237.
- [3] Mnich, Matthias, and Rene van Bevern. "Parameterized complexity of machine scheduling: 15 open problems." *Computers & Operations Research* (2018).
- [4] Blank, Julian, Kalyanmoy Deb, and Sanaz Mostaghim. "Solving the Bi-objective Traveling Thief Problem with Multi-objective Evolutionary Algorithms." *International Conference on Evolutionary Multi-Criterion Optimization*. Springer, Cham, 2017.
- [5] Bonyadi, Mohammad Reza, Zbigniew Michalewicz, and Luigi Barone. "The travelling thief problem: The first step in the transition from theoretical problems to realistic problems." *2013 IEEE Congress on Evolutionary Computation*. IEEE, 2013.
- [6] Bonyadi, Mohammad Reza, et al. "Evolutionary computation for multi-component problems: opportunities and future directions." *Optimization in Industry*. Springer, Cham, 2019. 13-30.
- [7] Martins, Marcella SR, et al. "HSEDA: a heuristic selection approach based on estimation of distribution algorithm for the travelling thief problem." *Proceedings of the Genetic and Evolutionary Computation Conference*. ACM, 2017.
- [8] Wagner, Markus. "Stealing items more efficiently with ants: a swarm intelligence approach to the travelling thief problem." *International Conference on Swarm Intelligence*. Springer, Cham, 2016.
- [9] Deb, Kalyanmoy, et al. "A fast and elitist multiobjective genetic algorithm: NSGA-II." *IEEE transactions on evolutionary computation* 6.2 (2002): 182-197.
- [10] Laszczyk, Maciej, and Paweł B. Myszkowski. "Improved selection in evolutionary multi-objective optimization of Multi-Skill Resource-Constrained project scheduling problem." *Information Sciences* 481 (2019): 412-431.
- [11] Laszczyk, Maciej, and Paweł B. Myszkowski. "Survey of quality measures for multi-objective optimization. Construction of complementary set of multi-objective quality measures." *Swarm and Evolutionary Computation* (2019).
- [12] Polyakovskiy, Sergey, et al. "A comprehensive benchmark set and heuristics for the traveling thief problem." *Proceedings of the 2014 Annual Conference on Genetic and Evolutionary Computation*. ACM, 2014.
- [13] Wu, Junhua, et al. "Exact approaches for the travelling thief problem." *Asia-Pacific Conference on Simulated Evolution and Learning*. Springer, Cham, 2017.
- [14] Vieira, Daniel KS, et al. "A genetic algorithm for multi-component optimization problems: the case of the travelling thief problem." *European Conference on Evolutionary Computation in Combinatorial Optimization*. Springer, Cham, 2017.
- [15] Wu, Junhua, et al. "Evolutionary computation plus dynamic programming for the bi-objective travelling thief problem." *Proceedings of the Genetic and Evolutionary Computation Conference*. ACM, 2018.
- [16] Whitley, L. Darrell, Timothy Starkweather, and D'Ann Fuquay. "Scheduling problems and traveling salesmen: The genetic edge recombination operator." *ICGA*. Vol. 89. 1989.
- [17] Marti, Luis, Eduardo Segredo, and Emma Hart. "Impact of selection methods on the diversity of many-objective Pareto set approximations." *Procedia Computer Science* 112 (2017): 844-853.
- [18] Yafrani, Mohamed, et al. "A hyperheuristic approach based on low-level heuristics for the travelling thief problem." *Genetic Programming and Evolvable Machines* 19.1-2 (2018): 121-150.
- [19] Cao, Yongtao, Byran J. Smucker, and Timothy J. Robinson. "On using the hypervolume indicator to compare Pareto fronts: Applications to multi-criteria optimal experimental design." *J. of Stat. Planning and Inference* 160 (2015): 60-74.
- [20] Durairaj, M., D. Sudharsun, and N. Swamynathan. "Analysis of process parameters in wire EDM with stainless steel using single objective Taguchi method and multi objective grey relational grade." *Procedia Engineering* 64 (2013): 868-877.
- [21] Lopez-Ibanez M., Paquete L., and Stutzle T. "Exploratory Analysis of Stochastic Local Search Algorithms in Biobjective Optimization", *Experimental Methods for the Analysis of Optimization Algorithms* (2010): 209-222.

# Urban Sound Classification using Long Short-Term Memory Neural Network

Iurii Lezhenin, Natalia Bogach

Institute of Computer Science and Technology  
Peter the Great St.Petersburg Polytechnic University  
St.Petersburg, 195251, Russia  
Email: {lezhenin, bogach}@kspt.icc.spbstu.ru

Evgeny Pyshkin

Software Engineering Lab  
University of Aizu  
Aizu-Wakamatsu, 965-8580, Japan  
Email: pyshe@u-aizu.ac.jp

**Abstract**—Environmental sound classification has received more attention in recent years. Analysis of environmental sounds is difficult because of its unstructured nature. However, the presence of strong spectro-temporal patterns makes the classification possible. Since LSTM neural networks are efficient at learning temporal dependencies we propose and examine a LSTM model for urban sound classification. The model is trained on magnitude mel-spectrograms extracted from UrbanSound8K dataset audio. The proposed network is evaluated using 5-fold cross-validation and compared with the baseline CNN. It is shown that the LSTM model outperforms a set of existing solutions and is more accurate and confident than the CNN.

**Index Terms**—environmental sound classification, long short-term memory, convolutional neural networks, UrbanSound8K dataset

## I. INTRODUCTION

AUDIO recognition algorithms are traditionally used for the tasks of speech and music signal processing. Meanwhile, the problems of environmental sound recognition and classification have received much attention in recent years. There are multiple applications already proposed in a big variety of industries, including surveillance [1], [2], audio scene recognition for robot navigation [3], acoustic monitoring of natural and artificial environment [4]–[6]. In digitally transformed society [7], soundscape models create a research perspective in smart city domain. City noise managing significantly contributes to a healthy and safe living environment in the big cities [8]. In travel centric systems, city sounds may enter the emerging solutions to develop and share journey experience [9], [10]. Assisting technologies for people with disabilities and, in particular, navigation systems for blind or visually impaired people effectively incorporate urban sound models [11].

Environmental sound analysis is more complex than speech and music processing because of unstructured nature of sounds. There are no meaningful sequences of elementary blocks like phonemes or strong stationary patterns such as melody or rhythm. However, environmental sounds may include strong spectro-temporal signatures. Thus, it is important to consider non-stationary aspects of signal and capture its variation in both time and frequency domains.

The classification of environmental sounds is often split into auditory scene classification and sound classification by

its source. But, both problems share the similar approaches. The methods used involve k-Nearest Neighbors (k-NN) algorithm, Support Vector Machine (SVM), Gaussian Mixture Model (GMM) and Hidden Markov Model (HMM) in combination with features engineered by signal processing techniques, e.g. Mel-Frequency Cepstral Coefficients (MFCC), Discrete Wavelet Transform (DWT) coefficients and Matching Pursuit (MP) features [12]–[14]. In contrast with described approaches, deep neural networks (DNN) allow to facilitate feature engineering keeping classification accuracy and even outperform the conventional solutions [15]. In particular, being able to capture spectro-temporal patterns from spectrogram-like input convolutional neural networks (CNN) have high performance [16]–[19]. Long short-term memory (LSTM) networks is the other type of neural network architectures that is exploited for sound classification [20], as well as the combinations of LSTM and CNN [21], [22].

LSTM networks are recurrent neural networks (RNN) that use the contextual information over long time intervals to map the input sequence to the output. LSTM network is a general solution, efficient at learning temporal dependencies. Its application is beneficial in a variety of tasks, such as phoneme classification [23], speech recognition [24] and speech synthesis [25]. LSTM network combined with CNN was also successfully used for video classification [26].

The applicability of LSTM for sound classification hasn't been fully investigated so far. In this paper we examine a LSTM model to improve understanding of its applicability specifically for urban sounds classification using UrbanSound8K dataset [27]. Table A1 in Appendix summarizes some of the existing solutions where models are evaluated on UrbanSound8K. The baseline accuracy of 70% was obtained with SVM processing mel-bands and MFCC statistically summarized across the time [27]. The unsupervised feature learning using Spherical K-Means (SKM) performed on PCA-whitened log-scaled mel-spectrograms allows to achieve 73.6% accuracy [28]. CNNs of different architectures trained on log-scaled mel-spectrogram frames provide 73% of accuracy and 79% with data augmentation [16], [17]. The LSTM based CRNN for urban sound classification demonstrates 79.06% accuracy using raw waveforms [22]. The accuracy of 93% was shown by GoogLeNet trained on combination

of mel-spectrogram, MFCC and Cross Recurrence Plot (CRP) images [18].

The paper is structured as follows: Section II describes the LSTM model studied and the experimental setup. In Section III we present and discuss our results, and, finally, in Section IV we conclude about the LSTM applicability for urban sound classification and provide directions for future work.

## II. METHOD

### A. Long-short term memory neural network model

LSTM neural network is a special kind of RNN, that doesn't suffer from vanishing gradient problem and is able to learn long-term dependencies. LSTM consists of a set of subnets, known as memory blocks. Each block includes the memory cell and three units: input, output and forget gates.

LSTM layer maps the input sequence  $X = (x_1, x_2, \dots, x_T)$  to the output sequence  $Y = (y_1, y_2, \dots, y_T)$  in according to the equations:

$$i_t = \text{sig}(W_{xi}x_t + W_{yi}y_{t-1} + b_i), \quad (1)$$

$$f_t = \text{sig}(W_{xf}x_t + W_{yf}y_{t-1} + b_f), \quad (2)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_{xc}x_t + W_{yc}y_{t-1} + b_c), \quad (3)$$

$$o_t = \text{sig}(W_{xo}x_t + W_{yo}y_{t-1} + b_o), \quad (4)$$

$$y_t = o_t \odot \tanh(c_t), \quad (5)$$

where  $c_t$  is the state of the memory cell and  $i_t, f_t, o_t$  are gate outputs at time  $t$ . The network weights  $W$  and biases  $b$  are tuned during learning to minimize the loss function. In case of a multi-layer structure the input of the next layer is the output of the previous one.

Our model for sound classification is composed of two LSTM layers followed by dense layer with *softmax* activation function. Though LSTM produces a sequence, only the last value is propagated to the output layer. The first two layers contain 128 and 64 units, the last layer has 10 units, one per sound class. To reduce overfitting dropout with a rate of 0.25 is applied to the output of the LSTM layers. For training *categorical cross-entropy* loss function is minimized using Adam optimizer. Because of long training time a full search of hyperparameters is infeasible, thus, the most promising combination was found using single fold evaluation.

The input of our model is magnitude mel-spectrogram with 128 bands, that covers a frequency range from 0 Hz to 22050 Hz. Spectrogram is evaluated at sample rate 44100 Hz using 1024 sample window and a hop size of the same width. The length of input sequence is variable and depends upon audio clip duration.

Among the examined variants the proposed model shows the best performance on input data normalized as follows:

$$\mu = \frac{1}{T} \frac{1}{N} \sum_{t=1}^T \sum_{n=1}^N x_t^{(n)}, \quad (6)$$

$$\sigma = \sqrt{\frac{1}{T} \frac{1}{N} \sum_{t=1}^T \sum_{n=1}^N (x_t^{(n)} - \mu)^2}, \quad (7)$$

$$X_{norm} = \frac{X - \mu}{\sigma}, \quad (8)$$

where  $X$  is the input sequence;  $x_t^{(n)}$  is the value of  $n$ -th feature at time  $t$ ;  $N$  is a number of features and  $T$  is a number of time steps. Normalization in both dimensions allows to keep spectro-temporal energy distribution pattern and eliminate the difference between the audio clips across the dataset in terms of linear distortion.

### B. Experimental setup

To evaluate the performance of proposed model we use UrbanSound8K dataset [27], that contains 8732 sound clips of up to 4 s in duration divided into 10 sound classes: air conditioner (AI), car horn (CA), children playing (CH), dog bark (DO), drilling (DR), engine idling (EN), gun shot (GU), jackhammer (JA), siren (SI), street music (ST).

Along with our model we run a baseline CNN [17]. CNN is composed of three convolutional layers followed by two dense layers. Both networks were trained on magnitude mel-spectrogram and CNN model indicated even better performance than was reported in [17] for log-scaled mel-spectrogram. We use a simplified validation algorithm for CNN: in contrast with [17], frame is being extracted from test sample at random, yet the CNN model holds the reported level of accuracy.

We randomly divide the dataset into 5 folds of the same size and carry out cross-validation to evaluate the networks performance. Models were trained on four folds and tested on the last one. The training duration is limited by 64 epochs. The train loss, train accuracy, test loss and test accuracy are saved for each epoch. The final accuracy is taken as the best validation accuracy achieved in the course of training.

Both models were implemented<sup>1</sup> with Keras, a high-level neural network API, written in Python. To resample the audio clips and extract the mel-spectrum we use the Librosa Python library.

## III. RESULTS AND DISCUSSION

Both models show the similar performance, their cross-validation results are presented in Fig. 1. While CNN provides 81.67% average accuracy, the proposed LSTM network

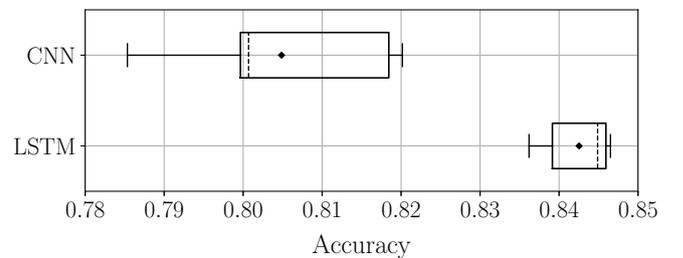


Fig. 1: Classification accuracy. Average accuracy is 80.48% and 84.25% for CNN and LSTM, respectively.

<sup>1</sup>Source code in Python available as Jupyter notebooks at <https://github.com/lezhenin/lstm-sound-classification-2019>

TABLE I: Per-class and averaged Precision, Recall and F1 score for CNN and LSTM.

		AI	CA	CH	DO	DR	EN	GU	JA	SI	ST	Macro-average
LSTM	Precision	0.80	0.82	0.78	0.86	0.87	0.88	0.93	0.89	0.90	0.75	0.85
	Recall	0.88	0.85	0.73	0.83	0.87	0.85	0.94	0.91	0.91	0.73	0.85
	F1	0.84	0.83	0.75	0.84	0.87	0.86	0.94	0.90	0.90	0.74	0.85
CNN	Precision	0.74	0.94	0.63	0.85	0.86	0.80	0.93	0.87	0.95	0.70	0.83
	Recall	0.83	0.79	0.71	0.80	0.81	0.84	0.89	0.84	0.83	0.73	0.81
	F1	0.78	0.86	0.67	0.83	0.83	0.82	0.91	0.85	0.88	0.71	0.82

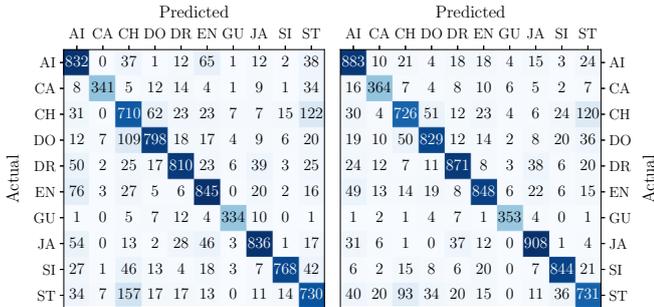


Fig. 2: Confusion matrices for CNN (left) and LSTM (right).

achieves 84.25%. The two models outperform the baseline methods. But LSTM demonstrates less accuracy distribution range and, thus, is more robust.

Confusion matrices obtained on test data during cross-validation is shown in Fig. 2. The same two pairs of classes demonstrate high confusion: street music vs. children playing and children playing vs. dog bark. These sounds may have complex time-frequency structure which impedes their accurate classification.

Precision, recall and F1 calculated for each class using confusion matrices are presented in Table I. LSTM shows slightly higher F1 score for each class, except car horn, and outperforms CNN in average. Also CNN may decrease recall to increase the overall accuracy, especially for unbalanced classes (e.g car horn and siren). Thus, LSTM performs better

keeping not only accuracy but recall and precision as well.

We compare training as accuracy and loss across epochs in Fig. 3. Both networks achieve the ultimate performance on test data approximately at 20-th epoch. Having almost equal accuracy the two models differ in their loss values. LSTM network shows a significantly smaller loss. It means LSTM is more confident in its predictions and has wider margins between classes. Thus, it is more robust.

The CNN holds accuracy and loss over train and test data. In contrast, LSTM model shows the better performance on train data. It doesn't fully generalize from train to unseen test data and memorizes the details that don't affect the overall performance. It may indicate that the model is redundant. Because of its recurrent structure the LSTM is more computationally intensive and prone to overfitting, although has less trainable parameters than CNN: 181K vs. 241K. So, it is highly probable that the model may be simplified without a significant performance degradation. Additional regularization techniques may also be beneficial.

#### IV. CONCLUSION

LSTM network that take magnitude mel-spectrograms was shown to be a reliable classifier in application for urban sounds. It provides the 84.25% of average accuracy and thus exceeds the majority of existing solutions. In comparison with baseline CNN trained on the same data LSTM has a little performance increase and is more confident.

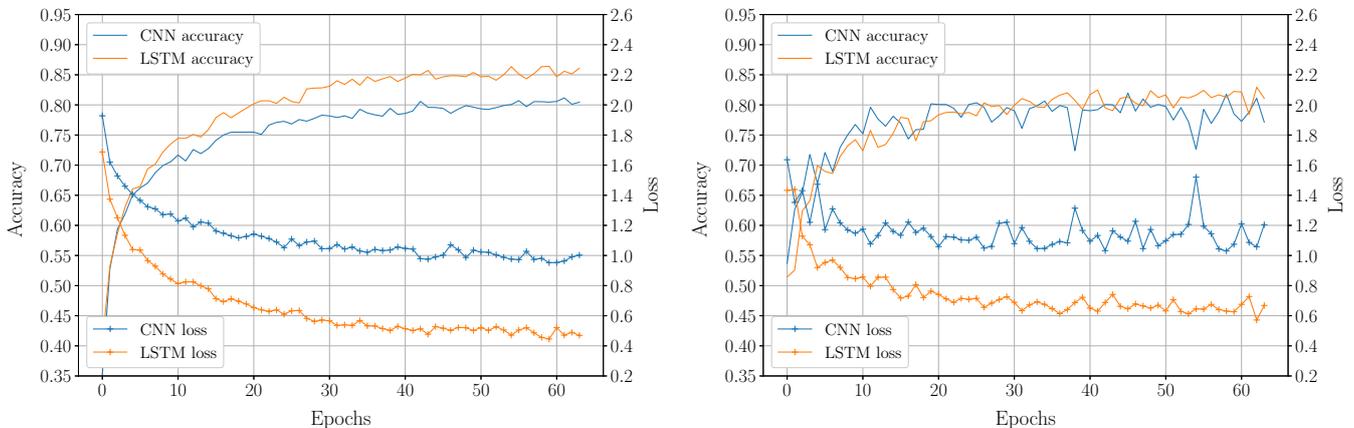


Fig. 3: Accuracy and loss evaluated on train data (left) and test data (right) during training.

The further study may develop towards the model simplification and regularization or involve new data not limited by urban setting.

#### APPENDIX

TABLE A1: Classification accuracy on UrbanSound8K dataset

Reference	Classifier	Features	Accuracy
[27]	SVM	mel-bands and MFCC	70%
[28]	SKM	PCA whitened mel-bands	73%
[16]	CNN	log mel-spectrogram	73%
[17]	CNN	log mel-spectrogram	73%
	CNN + aug		79%
[22]	CRNN	raw waveforms	79%
this paper	LSTM	mel-spectrogram	83%
[18]	CNN (GoogLeNet)	mel-spectrogram, MFCC, CRP images	93%

#### ACKNOWLEDGMENT

This work was partially supported by the grant 17K00509 of Japan Society for the Promotion of Science (JSPS).

#### REFERENCES

- [1] R. Radhakrishnan, A. Divakaran, and A. Smaragdis, "Audio analysis for surveillance applications," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2005. IEEE, 2005, pp. 158–161. [Online]. Available: <https://doi.org/10.1109/ASPAA.2005.1540194>
- [2] M. Cristani, M. Bicego, and V. Murino, "Audio-visual event recognition in surveillance video sequences," *IEEE Transactions on Multimedia*, vol. 9, no. 2, pp. 257–267, 2007. [Online]. Available: <https://doi.org/10.1109/TMM.2006.886263>
- [3] S. Chu, S. Narayanan, C.-C. J. Kuo, and M. J. Mataric, "Where am i? scene recognition for mobile robots using audio features," in *2006 IEEE International conference on multimedia and expo*. IEEE, 2006, pp. 885–888. [Online]. Available: <https://doi.org/10.1109/ICME.2006.262661>
- [4] R. Bardeli, D. Wolff, F. Kurth, M. Koch, K.-H. Tauchert, and K.-H. Frommolt, "Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring," *Pattern Recognition Letters*, vol. 31, no. 12, pp. 1524–1534, 2010. [Online]. Available: <https://doi.org/10.1016/j.patrec.2009.09.014>
- [5] C. Mydlarz, J. Salamon, and J. P. Bello, "The implementation of low-cost urban acoustic monitoring devices," *Applied Acoustics*, vol. 117, pp. 207–218, 2017. [Online]. Available: <https://doi.org/10.1016/j.apacoust.2016.06.010>
- [6] D. Steele, J. Krijnders, and C. Guastavino, "The sensor city initiative: cognitive sensors for soundscape transformations," *GIS Ostrava*, pp. 1–8, 2013.
- [7] V. Davidovski, "Exponential innovation through digital transformation," in *Proceedings of the 3rd International Conference on Applications in Information Technology*. ACM, 2018, pp. 3–5. [Online]. Available: <https://doi.org/10.1145/3274856.3274858>
- [8] F. Tappero, R. M. Alsina-Pagès, L. Duboc, and F. Alías, "Leveraging urban sounds: A commodity multi-microphone hardware approach for sound recognition," in *Multidisciplinary Digital Publishing Institute Proceedings*, vol. 4, no. 1, 2019, p. 55. [Online]. Available: <https://doi.org/10.3390/ecs5-5-05756>
- [9] E. Pyshkin, "Designing human-centric applications: Transdisciplinary connections with examples," in *2017 3rd IEEE International Conference on Cybernetics (CYBCONF)*. IEEE, 2017, pp. 1–6. [Online]. Available: <https://doi.org/10.1109/CYBCONF.2017.7985774>
- [10] E. Pyshkin and A. Kuznetsov, "Approaches for web search user interfaces-how to improve the search quality for various types of information," *JoC*, vol. 1, no. 1, pp. 1–8, 2010. [Online]. Available: <https://www.earticle.net/Article/A188181>
- [11] M. B. Dias, "Navpal: Technology solutions for enhancing urban navigation for blind travelers," *tech. report CMU-RI-TR-21, Robotics Institute, Carnegie Mellon University*, 2014.
- [12] S. Chu, S. Narayanan, and C.-C. J. Kuo, "Environmental sound recognition with time-frequency audio features," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 6, pp. 1142–1158, 2009. [Online]. Available: <https://doi.org/10.1109/TASL.2009.2017438>
- [13] S. Chachada and C.-C. J. Kuo, "Environmental sound recognition: A survey," vol. 3, 10 2013, pp. 1–9. [Online]. Available: <https://doi.org/10.1109/APSIPA.2013.6694338>
- [14] D. Giannoulis, E. Benetos, D. Stowell, M. Rossignol, M. Lagrange, and M. Plumbley, "Detection and classification of acoustic scenes and events: An ieeeaasp challenge," 10 2013, pp. 1–4. [Online]. Available: <https://doi.org/10.1109/WASPAA.2013.6701819>
- [15] Z. Koss, O. Toledo-Ronen, and M. Carmel, "Audio event classification using deep neural networks," in *Interspeech*, 2013, pp. 1482–1486.
- [16] K. J. Piczak, "Environmental sound classification with convolutional neural networks," in *2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 2015, pp. 1–6. [Online]. Available: <https://doi.org/10.1109/MLSP.2015.7324337>
- [17] J. Salamon and J. P. Bello, "Deep convolutional neural networks and data augmentation for environmental sound classification," *IEEE Signal Processing Letters*, vol. 24, no. 3, pp. 279–283, 2017. [Online]. Available: <https://doi.org/10.1109/LSP.2017.2657381>
- [18] V. Boddapati, A. Petef, J. Rasmusson, and L. Lundberg, "Classifying environmental sounds using image recognition networks," *Procedia computer science*, vol. 112, pp. 2048–2056, 2017. [Online]. Available: <https://doi.org/10.1016/j.procs.2017.08.250>
- [19] B. Zhu, K. Xu, D. Wang, L. Zhang, B. Li, and Y. Peng, "Environmental sound classification based on multi-temporal resolution convolutional neural network combining with multi-level features," in *Pacific Rim Conference on Multimedia*. Springer, 2018, pp. 528–537. [Online]. Available: [https://doi.org/10.1007/978-3-030-00767-6\\_49](https://doi.org/10.1007/978-3-030-00767-6_49)
- [20] Y. Wang, L. Neves, and F. Metzger, "Audio-based multimedia event detection using deep recurrent neural networks," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 2742–2746. [Online]. Available: <https://doi.org/10.1109/ICASSP.2016.7472176>
- [21] S. H. Bae, I. Choi, and N. S. Kim, "Acoustic scene classification using parallel combination of lstm and cnn," in *Proceedings of the Detection and Classification of Acoustic Scenes and Events 2016 Workshop (DCASE2016)*, 2016, pp. 11–15.
- [22] J. Sang, S. Park, and J. Lee, "Convolutional recurrent neural networks for urban sound classification using raw waveforms," in *2018 26th European Signal Processing Conference (EUSIPCO)*. IEEE, 2018, pp. 2444–2448. [Online]. Available: <https://doi.org/10.23919/EUSIPCO.2018.8553247>
- [23] A. Graves, S. Fernández, and J. Schmidhuber, "Bidirectional lstm networks for improved phoneme classification and recognition," in *International Conference on Artificial Neural Networks*. Springer, 2005, pp. 799–804. [Online]. Available: [https://doi.org/10.1007/11550907\\_126](https://doi.org/10.1007/11550907_126)
- [24] A. Graves, A.-r. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *2013 IEEE international conference on acoustics, speech and signal processing*. IEEE, 2013, pp. 6645–6649. [Online]. Available: <https://doi.org/10.1109/ICASSP.2013.6638947>
- [25] Y. Fan, Y. Qian, F.-L. Xie, and F. K. Soong, "Tts synthesis with bidirectional lstm based recurrent neural networks," in *Fifteenth Annual Conference of the International Speech Communication Association*, 2014.
- [26] J. Yue-Hei Ng, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, and G. Toderici, "Beyond short snippets: Deep networks for video classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 4694–4702. [Online]. Available: <https://doi.org/10.1109/CVPR.2015.7299101>
- [27] J. Salamon, C. Jacoby, and J. P. Bello, "A dataset and taxonomy for urban sound research," in *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 2014, pp. 1041–1044. [Online]. Available: <https://doi.org/10.1145/2647868.2655045>
- [28] J. Salamon and J. P. Bello, "Unsupervised feature learning for urban sound classification," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 171–175. [Online]. Available: <https://doi.org/10.1109/ICASSP.2015.7177954>

# Quantitative Impact of Label Noise on the Quality of Segmentation of Brain Tumors on MRI scans

Michał Marcinkiewicz

Netguru

ul. Wojskowa 6, 60-792 Poznań, Poland

Email: [michal.marcinkiewicz@netguru.com](mailto:michal.marcinkiewicz@netguru.com)

Grzegorz Mrukwa

Netguru

ul. Wojskowa 6, 60-792 Poznań, Poland,

Silesian University of Technology,

Data Mining Group

ul. Akademicka 16, 44-100 Gliwice

Email: [grzegorz.mrukwa@netguru.com](mailto:grzegorz.mrukwa@netguru.com)

**Abstract**—Over the last few years, deep learning has proven to be a great solution to many problems, such as image or text classification. Recently, deep learning-based solutions have outperformed humans on selected benchmark datasets, yielding a promising future for scientific and real-world applications. Training of deep learning models requires vast amounts of high quality data to achieve such supreme performance. In real-world scenarios, obtaining a large, coherent, and properly labeled dataset is a challenging task. This is especially true in medical applications, where high-quality data and annotations are scarce and the number of expert annotators is limited. In this paper, we investigate the impact of corrupted ground-truth masks on the performance of a neural network for a brain tumor segmentation task. Our findings suggest that a) the performance degrades about 8% less than it could be expected from simulations, b) a neural network learns the simulated biases of annotators, c) biases can be partially mitigated by using an inversely-biased dice loss function.

## I. INTRODUCTION

**T**HE HUMAN brain is proficient in recognizing patterns in a variety of domains: visual, auditory, etc. Its performance is always treated as the golden standard for the assessment and a level to beat using machine learning (ML) and deep learning (DL) models. As it stands, datasets are labeled by human annotators, with different levels of training, predispositions, and of course, also harbor their own biases, which have an impact on the quality of their annotations. Reducing the errors in datasets, also called label noise, calls for double- and triple-checking (usually done by different annotators), which requires a vast amount of work.

For example, in the classification of natural images – such as the ones included in the famous ImageNet dataset [1] – the human classification error rate was estimated at 5.1% by Russakovsky et al. [2]. However, the authors suggested that the labels provided by two human annotators did not exhibit strong overlap (one annotator's score was much lower – around 80%), and a significant amount of training was needed to achieve high-quality annotations.

The situation is even worse for more specialized domains, such as the diagnosis based on medical imaging, which requires years of training and experience. Moreover, due to the nature of the field, in some cases there is no clear way to classify a given observation correctly – studies showed that

medical diagnosis tests are not 100% accurate and cannot be considered the gold standard [3] [4]. This may be an effect of frequently occurring disagreements between medical experts interpreting test and imaging results [5] [6] [7] [8].

Image segmentation poses an even more severe problem. Reaching an agreement whether the object of interest is present in an image is relatively easy – what is challenging is to reach a consensus on its exact, pixel-wise location. In cases where more than one segmentation is available (which is seldom the case) there are multiple ways of handling such lack of consensus. An example of such method is the STAPLE algorithm [9], which automatically assigns confidence scores to each segmentation to merge multiple segmentations into one that is more accurate.

The presence of noise in annotations may even be more pronounced in real-world datasets, which are not carefully curated and annotated. Intuition tells us that training of a deep neural network (DNN) using a dataset with non-zero annotation noise can hurt the performance of a model, since loss function calculations provide "partially incorrect" gradients, which impair the learning process. Zhu et al. [10] investigated the effect of class label noise on the performance of a Decision Tree (DT) classifier in a classification task performed on various datasets. The study revealed that the performance of a DT classifier decays rapidly as the level of noise increases. Our recent investigation on a smaller scale (unpublished yet) revealed that classifiers based on DNNs can handle the rising amount of class label noise much better, even without applying any noise-filtering mechanisms.

## II. CONTRIBUTION

Another very important application of computer vision, besides image classification, is image segmentation. Image segmentation is often used in medical image processing, where segmentation masks provide a visual aid for physicians. In the future, it could become the first step of automatic or semi-automatic diagnosis processes. However, we must bear in mind that the annotations provided by DL-based models are heavily dependent on the quality of the data they were trained on. Our contribution presented in this paper is three-fold: a) we show the results of our investigation of the impact of various

levels of simulated noise in ground-truth segmentations on the performance of a DNN in brain tumor segmentation; b) a comparison of the DNN with a "perfect model", which learns perfectly the distribution of the simulated biases present in the data; c) the first results showing that an incorporation of bias into the loss function can partially combat a bias present in the data.

### III. DATA

In our study, we performed experiments on the BraTS2018 dataset [11], [12], [13], [14], which consists of MRI-DCE scans of 285 patients with diagnosed gliomas: 210 patients with high-grade glioblastomas, and 75 patients with low-grade gliomas. Each study was manually labeled by one to four expert readers. The data of each patient consists of 155 frames of size  $240 \times 240$  px, with four co-registered modalities: native pre-contrast T1-weighted (T1), post-contrast T1-weighted (T1c), T2-weighted (T2), and Fluid Attenuated Inversion Recovery (FLAIR). The scans were skull-stripped and interpolated to the same shape (155, 240, 240) with the voxel size of  $1 \text{ mm}^3$ . Each pixel was assigned one of the following four labels: healthy tissue (background), Gd-enhancing tumor (ET), peritumoral edema (ED), and necrotic and non-enhancing tumor core (NCR/NET) [12], [13], [14]. An example frame (T1c and T2) and the corresponding multiclass segmentation is shown in Fig 1. For the purpose of this work, all classes were merged into one – whole tumor (for a binary segmentation task).

Our pre-processing followed the methodology from the BraTS2018 competition presented in [15] – a volume-wise  $z$ -score normalization was applied to the brain region of each modality separately.

### IV. EXPERIMENT DETAILS

Our training was performed on a machine equipped with an Intel Core i7-7700 CPU, 64 GB RAM, and a NVIDIA GTX 1080 GPU. All experiments were performed with the PyTorch 1.0 framework in Python 3.6. In all experiments, we exploited a variation of U-net [16] with residual blocks [17] consisting of just under 1M parameters. The network consisted of 3 levels with 2 residual block on contracting path (CP) and expanding path (EP), for total of 12 residual blocks. Each residual block had 3 convolutional layers with 32, 48, and 64 filters on the first, second, and third level, respectively. The data from bridge connections (used between equivalent blocks on CP and EP) was concatenated in the channel dimension with the data coming from lower level, and a single convolutional layer was used to reduce the dimensionality. Parameters of the network were optimized by a SGD optimizer with the momentum of 0.9 and initial learning rate of 0.01. The learning rate was decreased by a factor of 5 after 10 and 16 epochs. The total length of training was 20 epochs, with batch size 14 (due to memory constraints). One epoch took around 22 minutes to train. For regularization we used weight decay of  $10^{-4}$ .

As the main objective function, we used the dice score (1), also called the  $f_1$ -score, which is a harmonic mean of *precision*

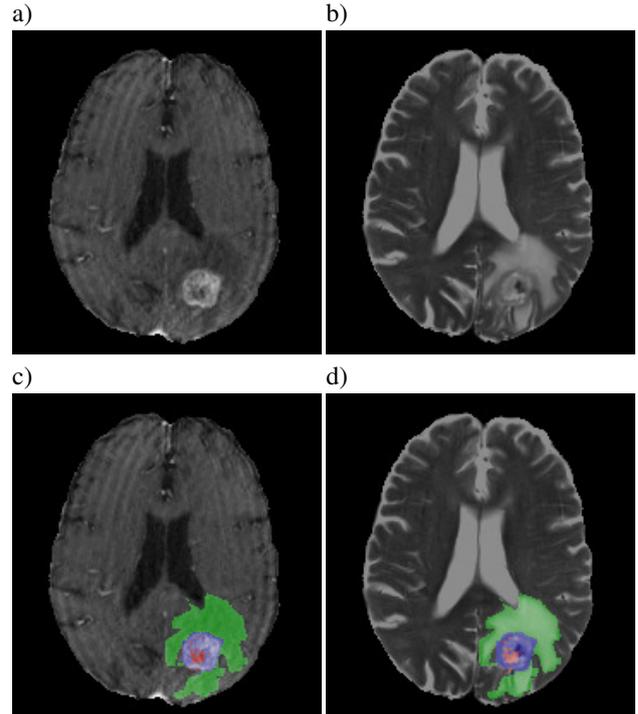


Fig. 1. Examples of images of BraTS2018 dataset in selected two modalities: a) T1c and b) T2. Their corresponding ground-truth segmentations are shown on panels c) and d), with three classes enhancing tumor (blue), peritumoral edema (green), and tumor core (red).

(positive predictive value), and *recall* (sensitivity). For the sake of differentiability we exploited its soft version (without thresholding). The pixel-wise dice score can be expressed as

$$Dice(p, t) = \frac{2 \sum_i p_i t_i + 1.0}{\sum_i p_i + \sum_i t_i + 1.0}, \quad (1)$$

where  $p_i \in [0, 1]$  is the predicted value at pixel  $i$ , and  $t_i \in \{0, 1\}$  is the target value of the same pixel, provided from the ground truth. To assure non-zero gradients and prevent division by zero, a smoothing factor of 1.0 was added to both the numerator and denominator. Since the popular DL frameworks are designed to minimize the objective function instead of maximizing it, we defined our loss function as

$$\mathcal{L}(p, t) = 1.0 - Dice(p, t). \quad (2)$$

The scores obtained on train and validation subsets were calculated for each frame, and then averaged; on the test subset, the scores were calculated volume-wise, which is a form of weighted-average with respect to the size of the ground-truth segmentation.

#### A. Data split

To validate our approach, we split the data into training, validation, and test subsets, containing 205, 40, and 40 data volumes, respectively. This allowed us to have 7 non-overlapping folds to perform cross-validation on. All the results presented are averaged over all folds.

### B. Simulated noise

To imitate sub-optimal segmentations, we assumed that even expert annotators can have their own biases, and their segmentations can have a noticeable variance due to human errors. We introduced biased noise to the train and validation subsets only, since we assumed that the test subset is of sufficiently high quality to be compared against. The bias-introduction routines were based on morphological operations applied to each frame with a binary mask using a  $3 \times 3$  structure one or more times. The morphological operations were incorporated in three ways:

- Dilate: simulates an annotator biased towards recall. The annotations produced tend to be over-segmented (the segmentations encapsulate more pixels than the true tumor), to be sure nothing important is missed. Since the tumor core is usually surrounded by the peritumoral edema, deciding exactly how far the tumor area reaches might be a non-trivial task.
- Erode: simulates an annotator biased towards precision. The annotations produced tend to be under-segmented, ensuring that only the tumor area is included. Because of that, some parts of the tumor can be omitted.
- Random: to simulate a random annotator or a mixture of annotators with different biases (either tending to over- or under-segment), we randomly assigned a dilation or an erosion operation for each frame in an accordingly sampled scale.

The number of iterations of morphological operations, denoted here as a scale of contamination, was sampled from a normal distribution  $\mathcal{N}(0, \sigma^2)$  with a few different values of variance  $\sigma^2 \in \{1, 2, 3, 4, 5\}$ . Since the number of iterations had to be a positive integer number, an absolute value of the number was taken, followed by an integer casting (the *floor* operation). The scale directly influenced the extent to which the original ground-truth mask was modified by a morphological operation (erosion / dilation) — it altered the relative change of size ( $\Delta S = S_{modified}/S_{original}$ ). If scale = 0, the ground-truth was fed into the network unchanged, meaning that  $\Delta S = 1$ . Some example effects of dilation and erosion operations applied to a selected frame of FLAIR modality are presented in Fig. 2 for scales  $\in \{0, 1, 3, 5\}$ . In panels (a) and (e), the ground-truth segmentation is unchanged. The dilation operation (top panels) increased the target segmentation size by 15%, 39%, and 61% for the scales of 1, 3, and 5, as shown in panels (b), (c), and (d), respectively. Erosion (bottom panels) decreased the target segmentation size by 14%, 39%, and 58% for the scales of 1, 3, and 5, as shown in panels (f), (g), and (h), respectively. It is worth pointing out that the magnitude of  $\Delta S$  depends strongly on the initial shape of a mask, thus morphological operations can introduce vastly different surface scaling factors.

## V. RESULTS

The average baseline test scores obtained by our model, without any modifications of the ground-truth segmentations,

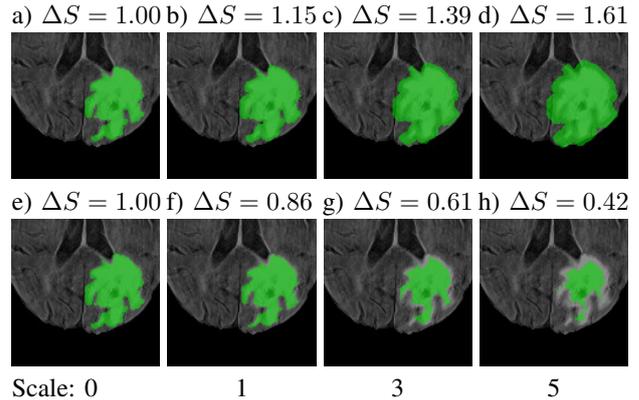


Fig. 2. Examples of biased noise of a binary mask overlaid on a FLAIR image selected from BraTS2018 dataset. Original mask is presented in panels a) and e), marked by the noise scale = 0 and relative change of size  $\Delta S = 1.00$ . Top row (panels b, c, and d) shows examples of dilation operation with scale  $\in \{1, 3, 5\}$ , which translates into  $\Delta S \in \{1.15, 1.39, 1.61\}$ . Bottom row (panels f, g, and h) shows examples of the erosion operation with the same scale, which translates into  $\Delta S \in \{0.86, 0.61, 0.42\}$ .

TABLE I  
BASELINE DICE, PRECISION, AND RECALL SCORES FOR OUR NETWORK TRAINED AND VALIDATED ON BRATS2018 DATASET FOR BINARY SEGMENTATION. MEAN AND STANDARD DEVIATION (STD) WERE CALCULATED OVER ALL FOLDS.

	Val Dice	Val Precision	Val Recall	Test Dice	Test Precision	Test Recall
Mean	0.896	0.906	0.880	0.872	0.902	0.863
Std	0.013	0.009	0.021	0.016	0.020	0.027

were 0.872, 0.902, and 0.863 for dice, precision, and recall, respectively. These values remained relatively consistent across all folds. The scores are comparable with some of the higher scores of the BraTS2018 challenge for the whole-tumor class on the training scoreboard. Unfortunately, since the challenge is over, we were not able to evaluate our results on the validation set or the test set, because the evaluation was carried out by the organizers. Following that, our results could not be compared with these submitted to the challenge by the participants. However, we would like to stress that multiclass segmentation (as in the BraTS2018 challenge) is a much more difficult task; the networks trained for the challenge might not have been optimized for binary segmentation, therefore there is no fair comparison between models trained for multiclass segmentation and our model. However, since our model reaches close to 0.9 of dice, precision, and recall, we are confident that it is good enough to act as a valid baseline.

The main results of this paper are presented in Fig. 3. The panels (a), (b), and (c), present the dice score, precision, and recall as a function of contamination scale, respectively. Solid lines represent the results obtained by our DNN for random (blue), dilation (orange), and erosion (green) contamination modes.

We performed a simulation of a "noise-robust" model – a model which has the same performance on noiseless data as

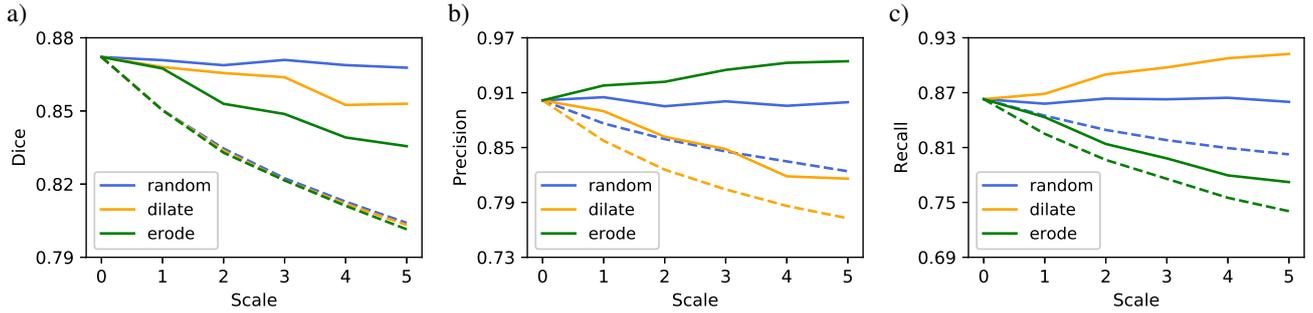


Fig. 3. Performance scores of a deep neural network trained using binary segmentation masks of BraTS2018 dataset with applied morphological noise simulation (erosion – orange, dilation – blue, or randomly chosen one – green) as a function of the scale of the noise. Panel (a) shows the dice score, panel (b) precision, and panel (c) recall.

our DNN, but also learns to mimic the noise-incorporation procedures, yielding the same performance at every scale assuming that the test set is noisy as well. Effectively, we altered the masks of each fold on each scale, and calculated all metrics against the original ground-truth segmentations. The procedure allowed us to verify how our DNN compares with the "noise-robust" model. The results for the simulated "noise-robust" model are presented with dashed lines in Fig. 3. The colors match the modes of the DNN.

#### A. Dice score

The dice score (Fig. 3a) shows a stable behavior for random noise, degrading slightly even for higher values of scale the dice score drops only about 0.004, from 0.872 to 0.868. In the case of dilation and erosion the drop is more significant, down to 0.853 and 0.836. The results obtained by our DNN for each mode are higher than the those obtained by the simulated learner by around 8% (random), 6% (dilate), and 6% (erode).

#### B. Precision

Random noise has a negligible effect on the precision score (Fig. 3b). Erosion biases data towards precision, which is reflected in the increase of the score for that mode, from 0.902 to 0.944. Dilation has an inverse effect – the score drops down to 0.816.

The precision score obtained by our DNN for each mode are higher than the those obtained by the simulated learner by around 8% (random), and 5% (dilate). Since erosion does not misplace any pixels, the noisy mask is contained completely within the original mask – the precision score is unaffected by such noise.

#### C. Recall

Random noise has similarly a negligible effect on the recall score (Fig. 3c). Dilation operation favors higher recall, which is reflected by the increase of the score for that mode from 0.863 to 0.912. Contrarily, the recall score for erosion drops down to 0.772.

The recall score obtained by our DNN for each mode are higher than the those obtained by the simulated learner by around 7% (random), and 4% (erode). Since dilation does not

misplace any pixels, the noisy mask encapsulates completely the original mask – the recall score is unaffected by such noise.

#### D. Reducing bias

We investigated whether biases present in the dataset could be proactively mitigated by altering the loss function (2). We tuned the relative weight of the precision and recall (parameter  $\beta$ ) of the dice score (1), generalizing it to the  $f_\beta$ -score, as in (3). This operation puts more attention of the loss function towards either recall (for  $\beta > 1$ ) or precision (for  $\beta < 1$ ), partially countering the biases present in the data. Particularly, for  $\lim_{\beta \rightarrow \infty} f_\beta$  yields *recall*, while  $\lim_{\beta \rightarrow 0} f_\beta$  *precision*.

$$f_\beta = (1 + \beta^2) \frac{\text{precision} \cdot \text{recall}}{\beta^2 \cdot \text{precision} + \text{recall}}, \quad (3)$$

where

$$\text{precision}(p, t) = \frac{\sum_i p_i t_i + 1.0}{\sum_i p_i t_i + \sum_i p_i (1 - t_i) + 1.0}, \quad (4)$$

and

$$\text{recall}(p, t) = \frac{\sum_i p_i t_i + 1.0}{\sum_i p_i t_i + \sum_i (1 - p_i) t_i + 1.0}. \quad (5)$$

To detect if the bias of the dataset could be mitigated and to what extent, we performed a gridsearch over multiple *beta* values  $\beta \in \{0.0, 0.2, \dots, 0.8\}$ . Lower values *beta* bias the loss function towards precision, so we biased the data towards recall by using dilation. The results of the gridsearch plotted as colormaps are shown in Fig. 4. The values for *beta* = 1.0 were already calculated (Fig. 3).

At no dilation (*scale* = 0) the dice score (Fig. 4a) decreases along with *beta*, which was expected as the network is no longer being trained to maximize the dice score directly. More importantly, the scores obtained for the values of *scale* and *beta* close to the anti-diagonal are visibly higher, especially for higher levels of noise, in comparison with the corresponding results for  $\beta = 1.0$ . For example, at  $\beta = 0.2$ , the network was able to quite consistently (for scale values  $\in \{3, 4, 5\}$ ) score around 1.5 percent higher than for the default dice-based loss function.

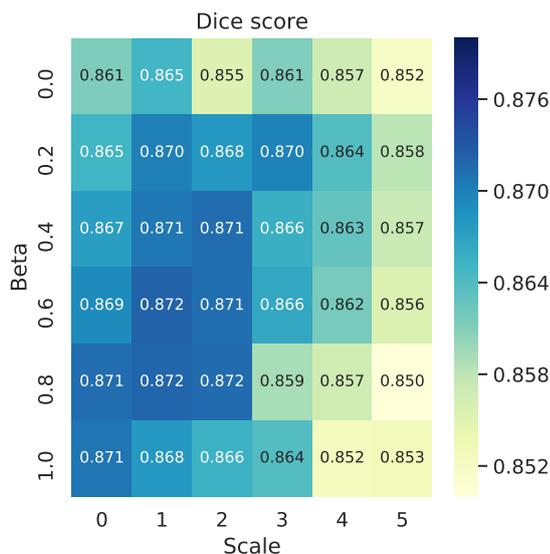


Fig. 4. Dice score of a deep neural network trained using binary segmentation masks of BraTS2018 dataset with applied morphological dilation as a function of the scale of the noise and a parameter  $\beta$ , representing the bias of the objective function towards precision (bias increases with decreasing  $\beta$ ).

Those results confirm that indeed the effect of bias in the dataset can be offset by incorporating an opposite bias in the objective function. Most likely, the optimal performance (obtained using unbiased dataset) cannot be restored completely, but, nevertheless, the gains are non-negligible. This puts the  $\beta$  parameter as a viable hyperparameter for optimizing the performance of a deep neural network in cases where there might be a bias present in a given dataset.

## VI. CONCLUSION

In this paper, we investigated the impact of simulated biases and variances of annotators—reflected in the under- or over-segmentation of binary mask they annotate—on the performance of a DNN trained on such modified image-mask pairs. We employed three types of simulated modifications of original ground-truth segmentation (which we called biased noise): erosion (simulating under-segmentation bias), dilation (simulating over-segmentation bias), and random, which employed randomly either erosion or dilation.

The results suggest that the performance of a DNN decays as the scale of contamination increases. The effect is rapid for both erosion and dilation, while it is slower (but steady) for the random contamination. This is because when training using under-segmented (eroded) segmentation masks, the DNN becomes biased towards precision, while using over-segmented (dilated) makes it biased towards recall. Both modes of contamination degrade the performance of a neural network significantly. However, for random contamination simulating a mixture of annotators with different biases, the decay of performance is less significant.

We also investigated whether the negative effect of a biased dataset on the training of a neural network could be reduced

by incorporating an opposite bias in the objective function. The results confirmed that both biases partially cancel each other, thus improving the performance. We suggest that the  $\beta$  parameter of the  $f_\beta$  score be considered as an important hyperparameter to search for during the optimization. Another option worth considering is to use multiple networks trained with different values of the  $\beta$  parameter in an ensemble. Such ensemble might improve the overall score via voting, just like an "ensemble" of expert annotators improve the score by improving the quality of ground-truth segmentations.

## REFERENCES

- [1] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, June 2009.
- [2] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. S. Bernstein, A. C. Berg, and L. Fei-Fei. Imagenet large scale visual recognition challenge. *Int J Comput Vis*, 115: 211, 2015.
- [3] L. Joseph, T. W. Gyorkos, and L. Coupal. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *Am. J. Epidemiol.*, 141(3):263–272, Feb 1995.
- [4] I. Bross. Misclassification in 2 x 2 tables. *Biometrics*, 10(4):478–486, 1954.
- [5] A. A. Bankier, D. Levine, E. F. Halpern, and H. Y. Kressel. Consensus interpretation in imaging research: is there a better way? *Radiology*, 257(1):14–17, Oct 2010.
- [6] W. R. Mower. Evaluating bias and variability in diagnostic test reports. *Ann Emerg Med*, 33(1):85–91, Jan 1999.
- [7] J. G. Jarvik and R. A. Deyo. Moderate versus mediocre: the reliability of spine MR data interpretations. *Radiology*, 250(1):15–17, Jan 2009.
- [8] J. A. Carrino, J. D. Lurie, A. N. Tosteson, T. D. Tosteson, E. J. Carragee, J. Kaiser, M. R. Grove, E. Blood, L. H. Pearson, J. N. Weinstein, and R. Herzog. Lumbar spine: reliability of MR imaging findings. *Radiology*, 250(1):161–170, Jan 2009.
- [9] S. K. Warfield, K. H. Zou, and W. M. Wells. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans Med Imaging*, 23(7):903–921, Jul 2004.
- [10] X. Zhu and X. Wu. Class noise vs. attribute noise: A quantitative study. *Artificial Intelligence Review*, 22(3):177–210, Nov 2004.
- [11] B. H. Menze et al. The multimodal brain tumor image segmentation benchmark (BraTS). *IEEE TMI*, 34(10):1993–2024, Oct 2015.
- [12] S. Bakas et al. Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific data*, 4:1–13, 9 2017.
- [13] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kirby, J. B. Freymann, K. F., and C. Davatzikos. Segmentation labels and radiomic features for the pre-operative scans of the TCGA-GBM collection, 2017. The Cancer Imaging Archive. <https://doi.org/10.7937/K9/TCIA.2017.KLXWJJ1Q>.
- [14] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kirby, J. B. Freymann, K. F., and C. Davatzikos. Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection, 2017. The Cancer Imaging Archive. <https://doi.org/10.7937/K9/TCIA.2017.GJQ7R0EF>.
- [15] M. Marcinkiewicz, J. Nalepa, P. R. Lorenzo, W. Dudzik, and G. Mrukwa. Automatic brain tumor segmentation using a two-stage multi-modal fcnn. In Alessandro Crimi, Spyridon Bakas, Hugo J. Kuijff, Farahani Keyvan, Mauricio Reyes, and Theo van Walsum, editors, *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, chapter 2, pages 13–24. Springer International Publishing, 2019.
- [16] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.
- [17] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015.



# Non-dominated Sorting Tournament Genetic Algorithm for Multi-Objective Travelling Salesman Problem

Paweł B. Myszkowski

Wrocław University of Science and Technology  
Faculty of Computer Science and Management  
ul. Ignacego Łukasiewicza 5, 50-371 Wrocław, Poland  
pawel.myszkowski@pwr.edu.pl

Maciej Laszczyk

Wrocław University of Science and Technology  
Faculty of Computer Science and Management  
ul. Ignacego Łukasiewicza 5, 50-371 Wrocław, Poland  
maciej.laszczyk@pwr.edu.pl

Kamil Dziadek

Wrocław University of Science and Technology  
Faculty of Computer Science and Management  
ul. Ignacego Łukasiewicza 5, 50-371 Wrocław, Poland  
220901@student.pwr.edu.pl

**Abstract**—A Travelling Salesman Problem (TSP) is an NP-hard combinatorial problem that is very important for many real-world applications. In this paper, it is shown, that proposed approach solves multi-objective TSP (mTSP) more effectively than other investigated methods, i.e. Non-dominated Sorting Genetic Algorithm II (NSGA-II). The proposed methods use rank and crowding distance (well-known from NSGA-II), combining those mechanisms in a novel, unique way: competing and co-evolving in the evolution process. The proposed modifications are investigated and verified by the benchmark mTSP instances, and results are compared to other methods.

## I. INTRODUCTION

A TRAVELLING Salesman Problem (TSP) is an NP-hard combinatorial optimization problem. The goal is to find a Hamiltonian cycle, that minimizes the sum of edge weights in a complete weighted graph [1]. Importance of TSP is accentuated by the fact, that it is a part of NP-complete class of problems [2].

A multi-objective Travelling Salesman Problem (mTSP) is an extension of TSP, where more than one objective is considered. It can be cost or time of the travel, the length of the route, etc. All objectives are optimized simultaneously [3]. An mTSP with two objectives is considered in this paper.

In multi-objective optimization, there is no prioritization of the objectives. Hence, to compare different solutions a dominance relation is used. A solution dominates another, if it has the value of at least one objective better, and value of no objectives worse than that solution. The goal of multi-objective optimization is to find all non-dominated solutions, a Pareto Front (PF). In practice, it is often not known whether found solutions comprise a true PF. Hence, the result of each method is called a PF approximation.

A Non-dominated Sorting Genetic Algorithm II (NSGA-II) [4], a classical multi-objective approach, uses two distinct mechanisms in its selection. The first one is the *rank*

comparison, which is based on the dominance relation, and aims to improve the convergence of the results. The second one is the *crowding distance*, which aims to increase the diversity of the results. However, a recent Non-dominated Sorting Tournament Genetic Algorithm (NTGA) [5] does not utilize the crowding distance at all. The authors show increased effectiveness of NTGA. This paper verifies the effectiveness of both *rank* and *crowding distance*. Two methods are presented that combine those mechanisms in a novel, unique way. One that uses them sequentially and forces competition between them. The other that utilizes two populations, where the mechanisms cooperate.

A set of experiments is designed to verify the quality of PF approximations generated by all methods. The Multi-Objective Evolutionary Algorithm integrating NSGA-II, SPEA2, and MOEA/D (MOEA/NSM) [12] is currently the best-known method for mTSP. Hence, it is used to compare the results. The results are evaluated by measuring convergence and diversity of the PF approximation and efficiency of the method. The set of Quality Measures (QMs) proposed in [6] is used. Moreover, visualizations of selected results are provided and a thorough theoretical analysis is presented.

The rest of the article is structured as follows. Section II contains the overview of existing work related to mTSP and multi-objective optimization. Section III provides a formal definition of the problem. All of the proposed approaches are described in section IV. Experiments and their results are presented in section V. The paper is concluded and additional remarks are given in section VI.

## II. RELATED WORK

A TSP is one of the most commonly researched problems. Many modifications to its original definition have been proposed. A TSP with asymmetric distances between the cities

[7], with multiple travelling salesmen [8], with stochastic travel times [9] or a vehicle routing problem [10].

Due to the NP-hard nature of mTSP, researchers often tackle it with metaheuristics. Genetic Algorithms are commonly used (e.g. [11], [12]). Authors of [14] and [15] have used an ant colony optimization methods. In [16] and [17] different memetic algorithms have been researched in the context of mTSP. It is also not uncommon to apply local search based methods [18], [19].

Researches often approach multi-objective optimization with genetic algorithms. They have proven to generate very high-quality PF approximations [20]. NSGA-II [4] is one of the most commonly used methods. It uses only a single population, where parents are forced to compete with children. It utilizes the rank and crowding distance of the individuals in the selection process and to truncate the population after every generation.

NTGA [5] is an extension of classical NSGA-II. The authors have shown its efficiency for a bi-objective scheduling problem – Multi-Skill Resource Constrained Project Scheduling Problem (MS-RCPSP). Modified selection in NTGA no longer utilizes the crowding distance. Instead, a clone elimination method is employed to maintain the diversity of the population. Additionally, the size of the tournament has been adjusted. Moreover, children are created in a new population and no longer have to compete with the parent population. NTGA uses also an archive that contains current approximation of PF. Such approach is a base of considerations in the given paper.

Recently created Multi-Objective Evolutionary Algorithm integrating NSGA-II, SPEA2, and MEA/D (MOEA/NSM) [12] has been successfully applied to mTSP. It uses the crowding distance, decomposition and Pareto strength. The solution space is explored using subpopulation tables, where each subpopulation contains the best results for a given aggregation of criteria. Every individual undergoes a crossover, mutation, and 2-opt optimization. At the end of each generation, subpopulations of SPEA2 and NSGA-II are updated. Rank and crowding distance mechanisms are considered. The authors show that MOEA/NSM outperforms all other methods and is the *state-of-the-art* population-based algorithm for mTSP. Hence, MOEA/NSM is used in this paper for comparison.

Initial multi-objective methods focused mostly on the convergence. However, recent research has shifted the focus onto the diversity [13]. This article tries to find the balance between the two. Two methods are proposed. One that switches the focus between convergence and diversity. The other that emphasizes both in parallel populations.

### III. PROBLEM

A TSP comprises of a set of  $m$  cities. In the problem a salesman must visit every city exactly once and return to the place where the travel started (initial city). The goal is to minimize the cost of travel of that route, given the cost of travel from city  $i$  to city  $j$ , is defined as  $c_{ij}$  and is part of the problem definition. TSP is equivalent to finding the minimum Hamiltonian cycle in a non-directed, weighted graph

[22], where nodes represent the cities, and weights represent the travel costs. Total cost of travel is calculated as the sum of edge weights and should be minimized. A symmetric TSP is considered in this paper, where  $c_{ij} = c_{ji}$  for all cities  $i, j \in \{0, 1, \dots, m-1\}$

In Multi-objective Travelling Salesman Problem (mTSP) multiple aspects of the route are evaluated [12]. It could be cost, time, length or risk of travel. The problem with  $n$  criteria and  $m$  cities is represented by  $n$  weighted graphs. For each  $k \in 1, \dots, n$ , graph  $G_k$  is a weighted graph, that represent  $k$ -th criterion. The edge weight between cities  $i$  and  $j$  in graph  $G_k$  is represented by  $c_{ij}^{(k)}$ . In this paper two criteria are considered.

### IV. EVOLUTIONARY METHODS

This section contains the description of all the investigated methods. First, definitions of important terms related to the work are given. Next, parts that are common for each method are described. Then, reference methods are presented. Finally, two novel modifications are described.

#### A. Definitions of Terms

1) *Dominance Relation*: The comparison of multi-objective solutions is done with the dominance relation. Let  $z, z'$  be two points in the multi-objective solution space.  $z$  dominates  $z'$  when both Eq.1 and Eq.2 are satisfied:

$$\forall_{k=1}^n f_k(z) \leq f_k(z') \quad (1)$$

$$\exists_{k=1}^n f_k(z) < f_k(z') \quad (2)$$

Where  $n$  is the number of criteria,  $f_k$  is the objective function of  $k$ -th criterion.

2) *Pareto Front*: A set of all non-dominated solutions is called a Pareto Front (PF). Since the set of globally non-dominated points is not known, all methods create an approximation of PF.

#### B. Representation

The representation of an individual in genetic algorithm defines how a genome represents the solution in a given problem. It also determines the use of genetic operators. All methods in this paper use the same representation.

An individual, for the problem with  $m$  cities, is represented by the permutation vector  $z = (m_1, m_2, \dots, m_m)$ . Each gene is the number of the next city on the route. In TSP the full route must end on the same city that it started. Hence, in the calculation of the objective functions the cost of travel between  $m_m$  and  $m_1$  must also be considered.

#### C. Initial Population

The first step of a genetic algorithm is the generation of an initial population. A random initialization is used. Every individual is initialized with a random permutation of all  $m$  cities. An additional mechanism enforces the uniqueness of all generated genotypes.

#### D. Genetic Operators

This chapter contains the description of both crossover, and mutation, which allow for exploitation, and exploration of the solution space.

1) *Crossover*: Crossover operator is responsible for the exploitation of space [23]. In the process two parent individuals are used to create two children individuals. In all methods, crossover is performed with a given probability ( $P_x$  parameter). In case of no crossover, parent genomes are copied over to the children individuals.

An Order Crossover (OX) has been selected [24]. It tends to retain the relative order of the genes and has been proven to work well for the ordering problems [25]. First, a part of the route is copied from the first parent, and then the rest of the route is reconstructed based on the genomes of the second parent. The part to copy is selected by randomly choosing two cut-points of the chromosome. The part between those two points is selected and copied to the child individual (in the same place of the genome). The remaining genes are filled, starting with the second cut-point, from the second parent. The order is maintained and already existing cities are skipped.

For example, given two parent individuals  $p_1$  and  $p_2$ :

$$\begin{aligned} p_1 &= (3 \ 2 \ 1 \mid 8 \ 4 \ 6 \ 7 \mid 5 \ 9), \\ p_2 &= (2 \ 3 \ 6 \mid 5 \ 8 \ 4 \ 1 \mid 9 \ 7). \end{aligned}$$

The first child  $c_1$  is:

$$c_1 = (3 \ 5 \ 1 \mid 8 \ 4 \ 6 \ 7 \mid 9 \ 2).$$

The second child  $c_2$  is generated by swapping the roles of two parents in the crossover process:

$$c_2 = (2 \ 6 \ 7 \mid 5 \ 8 \ 4 \ 1 \mid 9 \ 3).$$

2) *Mutation*: Mutation introduces a random perturbation in the genome and is responsible for exploration of the solution space [1]. It introduces one or more small changes within the genome with given probability  $P_m$ . The parameter is a probability of a mutation of a single individual.

An *Inversion* mutation has been selected. It performs an inversion of a randomly selected sequence of genes. The sequence is selected by randomly choosing two cut-points within the genome. All genes between those points are inverted.

For example, let's consider a parent  $p$  with the following genome and selected cut-points:

$$p = (3 \ 2 \ 1 \mid 8 \ 4 \ 6 \ 7 \mid 5 \ 9).$$

Mutation would result in the following genome  $c$ :

$$c = (3 \ 2 \ 1 \mid 7 \ 6 \ 4 \ 8 \mid 5 \ 9).$$

#### E. Selection

Selection operator is used to provide parent individuals for the genetic operators. It pressures the evolutionary process towards the desired results. In the case of multi-objective optimization it is important to find the PF approximation close to the true PF, but also to promote the diversity of the population. However, selection must also allow for the weaker

individuals in order to avoid local optima. In multi-objective optimization the selection is based on the rank and crowding distance [4].

The *rank* is calculated based on the dominance relation. First, all non-dominated individuals within the population gain rank 1. Then those individuals are exempt from further calculations. Rank 2 is assigned to non-dominated individuals from the remaining individuals. The process is iteratively repeated, until every individual has a rank assigned. Higher rank means that the individual is closer to the true PF.

The *crowding distance* is calculated based on the distance to other individuals. It is a volume of the largest cube that contains only that individual. A larger value means that there are fewer individuals in that part of the space.

Researched methods use a *tournament* selection. First, given number of individuals is randomly drawn from the population. They are compared according to given selection operators. The best individual, according to the operators, is returned. NSGA-II originally uses a tournament selection with two individuals, while NTGA allows for higher values of the tournament size. Moreover, in selection method NTGA uses an *archive* that contains all non-dominated individuals found in a given evolution process.

#### F. Evolutionary Process

The same evolutionary process is used in all methods. It is used to generate a new population  $P_{next}$ , from the current population  $P_{current}$ . It also includes a clone prevention method and archive usage introduced by NTGA. The process is described in pseudocode 1.

---

#### Algorithm 1 Pseudocode of the evolutionary process

---

```

1:  $P_{next} \leftarrow \emptyset$ 
2: while  $|P_{next}| < populationSize$  do
3:    $parents \leftarrow select(P_{current} \cup archive, operators)$ 
4:    $children \leftarrow mutate(crossover(parents))$ 
5:   while  $P_{next}$  contains  $children$  do
6:      $children \leftarrow mutate(children)$ 
7:   end while
8:    $evaluate(children)$ 
9:    $P_{next} \leftarrow P_{next} \cup children$ 
10: end while
11: return  $P_{next}$ 

```

---

In line 1 of Pseudocode 1, the next population is initialized to an empty collection. The loop between lines 2 and 10 performs the evolution until the next population reaches the desirable size. In line 3, parents are selected from the current population. NSGA-II does not use archive, but NTGA does (see line 3) in the selection method. The comparison is performed based on selected operators. Then, the children are created by performing crossover and mutation on the parents in line 4. Lines 5 to 7 describe the clone prevention mechanism. As long as the children already exist in the next population, they are mutated. Eventually, the children are evaluated in

line 8, and added to the next population in line 9. The next population is returned in line 11.

### G. Switch Non-Dominated Tournament Genetic Algorithm

The proposed method switches selections (*crowding distance* and *rank*) to obtain an evenly distributed PF approximation with high spread. Switch Non-dominated Tournament Genetic Algorithm (sNTGA) is based on a recent NTGA [5]. It switches between two competing selection operators - primary  $S_{primaryOperator}$  and a temporary one  $S_{switchOperator}$ . The former is used to obtain a diverse PF approximation, while the latter to also guarantee the convergence of the approximation. Both operators work in turns, and they promote different solutions, which allows for a better exploration of the solution space. Additionally, switch of the operators makes it easier to escape local optima. In consequence, improved solutions can be achieved. The time-frame, in which the operators work is defined by the number of births and the following parameters.

- $S_{delay}$  - number of births, after which the temporary operators is switched on. At the very beginning of the evolution, the switch is not necessary, as the population is still diverse, and not yet converged.
- $S_{each}$  - parameter determining a single cycle of the operators
- $S_{duration}$  - number of births after which  $S_{switchOperator}$  should be switch off and  $S_{primaryOperator}$  should be switched back on. It should always be lesser than  $S_{each}$ .

Pseudocode 2 describes the sNTGA.

---

#### Algorithm 2 Pseudocode of sNTGA

---

```

1:  $archive \leftarrow \emptyset$ 
2:  $P_{current} \leftarrow generateInitialPopulation()$ 
3:  $evaluate(P_{current})$ 
4: while  $stoppingCriteria()$  do
5:    $nonDominatedSorting(P_{current})$ 
6:    $crowdingDistanceAssignment(P_{current})$ 
7:    $updateArchive(P_{current})$ 
8:    $operator \leftarrow selectionOperator()$ 
9:    $P_{current} \leftarrow evolve(P_{current} \cup archive, operator)$ 
10: end while
11:  $updateArchive(P_{current})$ 
12: return  $archive$ 

```

---

An empty archive is initialized in line 1. In line 2, the current population is randomly initialized. It is then evaluated in line 3. The loop between lines 4 and 10 runs until the stopping criteria is reached. In the article it has been set to the given number of births. In line 5, the non-dominated sorting is performed and the crowding distance is assigned in line 6. The archive is updated in line 7, by adding all non-dominated individuals, and removing those that became dominated. In line 8 a selection operator is determined (described in pseudocode 3). Then the current population is evolved (described in pseudocode 1), based on the current individuals, the archive, and selection operators. Finally, the archive is again updated

in line 11 and it is returned in line 12, where it contains the PF approximation.

---

#### Algorithm 3 Pseudocode of $selectionOperator$ in sNTGA

---

```

1:  $switchBirthsCount \leftarrow currentBirthsCount - S_{delay}$ 
2: if  $switchBirthsCount < 0$  then
3:    $operator \leftarrow S_{primaryOperator}$ 
4: else if  $switchBirthsCount \bmod S_{each} < S_{duration}$  then
5:    $operator \leftarrow S_{switchOperator}$ 
6: else
7:    $operator \leftarrow S_{primaryOperator}$ 
8: end if
9: return  $operator$ 

```

---

To determine the current selection operators for sNTGA, the moment, at which operator should change is calculated in line 1. A  $switchBirthsCount$  variable is used. It is calculated as the current number of births minus the delay parameter. If that number is smaller than 0 (check in line 2) then the primary operator is selected in line 3. Otherwise, a check is performed to verify, which operator should be used, in line 4. If not enough births have happened, then the temporary operator is used in line 5, otherwise primary operator is used in line 7. The selected operator is returned in line 9.

Researched sNTGA uses the *crowding distance* as the primary operator and *rank* as the temporary operator.

### H. Co-Evolutionary Non-Dominated Tournament Genetic Algorithm

Both *rank* and *crowding distance* operator can be successfully used in multi-objective optimization methods. However, since both operators work on the same population, one operator might diminish the effect of the other operator. Hence, a Co-evolutionary Non-Dominated Tournament Genetic Algorithm (cNTGA) is proposed. The main motivation in cNTGA is to enforce cooperation of selection methods by operating on two separate populations connected in the evaluation process. Thus co-evolution mechanism is applied.

A cNTGA utilizes two populations with different selection operators. The exchange of information is possible due to the use of the same archive. The archive is used in the evaluation and selection process of both populations. Additionally, at the end of each generation, individuals of both populations are added to the archive. The method requires only one additional parameter  $K_{RANK}$  - it defines the percentage of the initial population size, that should be assigned to the population that uses *rank* operator. cNTGA is described in pseudocode 4.

**Algorithm 4** Pseudocode of cNTGA

---

```

1: archive ← ∅
2:  $P_{RANK}, P_{CD} \leftarrow generateInitialPopulation()$ 
3:  $evaluate(P_{RANK})$ 
4:  $evaluate(P_{CD})$ 
5: while  $stoppingCriteria()$  do
6:    $nonDominatedSorting(P_{RANK})$ 
7:    $nonDominatedSorting(P_{CD})$ 
8:    $crowdingDistanceAssignment(P_{CD})$ 
9:    $updateArchive(P_{RANK} \cup P_{CD})$ 
10:   $P_{RANK} \leftarrow evolve(P_{RANK} \cup archive, \geq_{RANK})$ 
11:   $P_{CD} \leftarrow evolve(P_{CD} \cup archive, \geq_{CD})$ 
12: end while
13:  $updateArchive(P_{RANK} \cup P_{CD})$ 
14: return archive

```

---

An empty archive is initialized in line 1. Then two populations  $P_{RANK}$  and  $P_{CD}$  are randomly initialized in line 2. Then, they are evaluated in lines 3 and 4 respectively. The loop between lines 5 and 12 runs until the stopping criteria is reached. In the paper, the number of births is used. In lines 6 and 7, the populations are sorted according to the appropriate operator. The crowding distances are assigned for population  $P_{CD}$  in line 8. Then, the archive is updated in line 9, with the individuals from both populations. The populations are evolved (pseudocode 1) in line 10 and 11 using  $\geq_{RANK}$  and  $\geq_{CD}$  operators. The archive is once again updated in line 13. Finally, the archive containing the PF approximation is returned in line 14.

In cNTGA, co-evolution significantly reduces the number of parameters (comparing to sNTGA) and results in an easier investigation process. Finally, the method decides which selection operator is more “useful” in the given problem, not the researcher in the tuning process.

*I. Reference Methods*

The researched methods are compared to three selected, reference methods. First, a NSGA-II [4] has been selected, as it is the most common method in the literature. NTGA [5] is selected, because approaches described in this paper are based on it. Finally, MOEA/NSM [12] is used as the current best *state-of-the-art* method.

## V. EXPERIMENTS AND RESULTS

This section presents experimental procedure to verify effectiveness of the proposed methods (sNTGA and cNTGA) by empirically comparing results to other reference methods. Thus, used mTSP instances are presented, Quality Measure for multi-objective optimization and methods setup are given. Finally, results and selected visualizations are presented.

*A. Data Instances*

Data instances used in the research are commonly used in literature, e.g. 9 instances *euclid\*\*\** from TSPLIB (e.g. [21]) and {kroAB100, kroAB200} generated from DIMACS code. Instances differ in number of cities, which affects the complexity and size of the solution landscape.

*B. Quality Measures*

To evaluate the results, the QMs proposed in [6] are used. This section contains their description along with the description of reference points required for their calculation.

1) *Euclidean Distance*: Euclidean distance (*ED*) is the average distance between every point on the PF approximation and a so called Perfect Point. Where the Perfect Point comprises of the best values of all objectives. ED can be formally defined by equation 3.

$$ED(PF) = \frac{\sum_{i=1}^{|PF|} d_i}{|PF|} \quad (3)$$

Where  $PF$  is the Pareto Front,  $d_i$  is the distance from the  $i$ 'th point to the Perfect Point.

ED measures the convergence of the PF approximation and should be minimized.

2) *Hypervolume*: Hypervolume (*HV*) is the volume of hypercube defined by the PF approximation and the Nadir Point. Where the Nadir Point comprises of the worst values of all objectives. HV can be formally defined by equation 4.

$$HV(PF) = \Lambda\left(\bigcup_{s \in PF} \{s' | s \prec s' \prec s^{nadir}\}\right) \quad (4)$$

Where  $PF$  is an approximation of PF.  $s$  is the point of approximated PF.  $s^{nadir}$  is a *Nadir Point*.  $\Lambda$  is a Lebesgue measure, which generalizes the a volume.  $\prec$  is a domination relation.

Hypervolume is a measure of spread, but is also influenced by the convergence of the PF approximation. It should be maximized.

3) *Pareto Front Size*: Pareto Front Size (*PFS*) is the number of points on the PF approximation. It is the measure of diversity and should be maximized.

4) *Spacing*: Spacing (*S*) is the average distance between the consecutive points on the PF approximation. S can be formally defined by equation 5.

$$S(PF) = \sqrt{\frac{1}{|PF|} \sum_{i=1}^{|PF|} (d_i - \bar{d})^2} \quad (5)$$

Where  $PF$  is the approximation of the PF.  $d_i$  is the distance from the  $i - th$  point the next consecutive point.

S is the measure of uniformity and should be minimized.

5) *Ratio of Non-dominated Individuals*: Ratio of Non-dominated Individuals (*RNI*) is the value of *PFS* divided by the number of births. It measures efficiency of the method and should be maximized.

6) *Purity*: Purity is used for a direct comparison of two PF approximations. It is the number of points that remain non-dominated, when combined with the PF approximation from another method. Purity should be maximized.

### C. Parameters

Each method is tuned experimentally (like cNTGA, sNTGA) or optimal configuration has been used based on publications (e.g. MOEA/NSM [21] or NTGA [5]). Each method is limited by number of *births* (a number of all visited points). All configurations are presented in Tab. I.

For all investigated methods this value is given according to number of cities as follows:

- 100 cities – 10 mln births,
- 200 cities – 13 mln births,
- 300 cities – 16 mln births,
- 500 cities – 22 mln births,

Such *births* limitations are connected to the size of landscape (number of cities) and limit of computational time required by experimental procedure.

All investigated methods (NTGA, sNTGA, cNTGA and NSGA-II) have been implemented in Java using standard libraries. An exception is MOEA/NSM, where authors of [21] code<sup>1</sup> has been used.

### D. Experiments

Averaged results from 50 runs of 5 methods for 11 instances are presented in Table II. The comparison shows that in each case the best values of *RNI* have been achieved by cNTGA or sNTGA. Proposed methods also outperform others in *PFS* context – they give 2–3 times larger number of points in approx. PF.

NSGA-II created interesting results – in 6/11 cases the best *ED* and *S* values have been achieved – it is connected to the fact, that this method gives very narrow PF approx. that is located in the “centre” (see Fig.1 or Fig.2). Other cases (5/11) are occupied by NTGA, what confirms that *ED* prefers methods that focus on PF “centre”.

In opposite cases (4/11) better *S* values were obtained by cNTGA with much large *PFS* value. Almost every time (9/11 cases) the best value of *HV* was achieved by MOEA/NSM.

All results have been averaged and presented in Table III. Two proposed methods (sNTGA and cNTGA) compete successfully with MOEA/NSM. In case of *ED*, all mentioned methods give results that are almost the same. The Wilcoxon signed-rank test showed that results of these three methods are not statistically different.

Results presented in Table III show that MOEA/NSM gives better approx. PF in *HV* context. The difference between sNTGA and MOEA/NSM results is statistically significant – Wilcoxon signed-rank test ( $W_{0.05} = 66 > W_c = 13$ ) confirmed that.

For other QM’s, two proposed methods – cNTGA and sNTGA – give results (*ED*, *RNI* and *S*) that are not statistically different. The exception is *HV*, where sNTGA outperforms cNTGA and it is statistically significant ( $W_{0.05} = 65 > W_c = 13$ ). However, both methods outperform MOEA/NSM. E.g. comparing results for cNTGA and MOEA/NSM Wilcoxon signed-rank test gives: *PFS* ( $W_{0.05}$

$= 66 > W_c = 13$ ), *RNI* ( $W_{0.05} = 64 > W_c = 13$ ) and *S* ( $W_{0.05} = 66 > W_c = 13$ ).

### E. Visualizations and approx. PF analysis

To visualize approximations of PF, selected graphs have been prepared. They present the “averaged” PF – modified version of empirical attainment function (EAF) [26]. Each graph contains data of PF from 50 independent runs of selected method.

For smaller instances (i.e. euclidAB100) the results of 5 investigated methods are very similar. Hence the visualization has been omitted. Only two methods – NSGA-II and NTGA – give worse results: approx. PF is dominated by others and have larger standard deviation.

Graph presented on Fig.1 shows results for more complex instance, euclidAB300. Methods cNTGA, sNTGA and MOEA/NSM compete successfully. NTGA gives the worst solution. Quite interesting are results of NSGA-II – approx. PF is too short but focused in central region of PF.

The more difficult instance (see Fig.2, euclidAB500) suggests that MOEA/NSM is effective in the central area of PF – QM’s *S* and *HV* confirm that. However, in other areas, sNTGA and cNTGA can compete and give very good results – a large number of points. Moreover, PF approximation created by NSGA-II is very short but of very high quality and focused in the central area.

To get a more detailed image of results **Purity** measure have been (see Tab. IV) used to compare gained approx. PF, where methods are compared in pairs. Results show that it is quite difficult to select the method with the best results. Fig.??- Fig.1 showed that MOEA/NSM gives better results in “centre” region, cNTGA wins in other regions. Using Purity measure, it gives domination of MOEA/NSM in 53.9% but cNTGA dominates in 54.3%. Detailed analysis of the data shows that cNTGA “wins” in instances with 500 cities and fails in 300 cities – it can be a suggestion that *births* limits cNTGA to much and such aspect should be investigated more carefully in further research. It is worth mentioning that cNTGA and sNTGA give significantly larger PF approx. (see *PFS* values in Tab.II or Tab.III) what can disturb a little interpretation of Purity values.

Another conclusion gained from Tab.IV is that cNTGA outperforms results of sNTGA in each instance – approx. PF “dominates” in 62%. Moreover, sNTGA does not compete with MOEA/NSM so successfully, and is “dominated” in 59.8%.

### F. Summary

cNTGA and sNTGA methods are effective and can compete with MOEA/NSM results (e.g. better *PFS* and *S* values). However, there is a strong need to focus the methods in central area of PF. Moreover, proposed methods have better efficiency than MOEA/NSM – the *RNI* confirmed that. The last but not least, is the conceptual aspect – cNTGA and sNTGA are methods that have only few parameters, are easy to understand and tune.

<sup>1</sup>MOEA/NSM code: <https://github.com/MOEA-NSM/moea-nsm>

TABLE I  
CONFIGURATIONS FOR INVESTIGATED METHODS

	NSGAI	MOEA/NSM	NTGA	cNTGA	sNTGA
mutation ( $P_m$ )	Swap, 10%	Swap, 10%	Inversion, 0.5%	Inversion, 0.25%	Inversion, 0.5%
crossover ( $P_c$ )	OX, 80%	OX, 80%	OX, 40%	OX, 70%	OX, 60%
tournament size ( $T_{size}$ )	2	2	25	25	25
pop_size ( $K$ )	500	100	500	500	500
additional parameters		$P_{obj}^1, P_{obj}^1$ 20 $P_{pon}$ 40 $M_t$ 100 $N_t, S_t$ 50 $P_{2OPT}$ 10%		$K_{RANK}$ 60%	$S_{delay}$ 4 000 000 $S_{each}$ 1 000 000 $S_{duration}$ 750 000

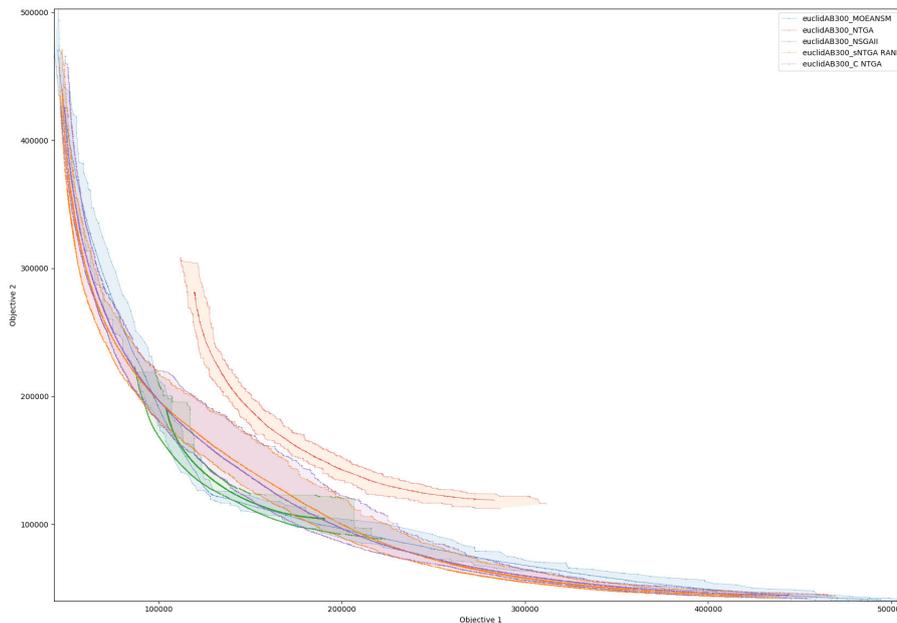


Fig. 1. Comparison of average approx. Pareto Fronts for data instance euclidAB300

## VI. CONCLUSIONS AND FUTURE WORK

This article introduces two new methods based on recent NTGA. The first one is sNTGA, which uses the rank and crowding distance operators sequentially. The second one - cNTGA utilizes two subpopulations, where one of them uses rank, and the other one uses crowding distance in their respective selection methods. Both subpopulations cooperate by performing the selection on a shared archive of non-dominated individuals.

The methods are compared to the *state-of-the-art* population-based MOEA/NSM. Proposed methods require fewer parameters and it is argued that they are less complex. Additionally, the results are evaluated with a set of QMs on the selected instances of TSP.

Another promising trend in literature is specialization and hybridization of metaheuristics. Proposed methods could be

extended by some local search techniques like 2-opt or/and genetic operators that are considered to be more effective in TSP, like Edge Crossover Operator [27].

It is shown that MOEA/NSM achieves better results in the 'centre' of the PF approximation. A different selection method that would emphasize the convergence in that area could be beneficial. Also, research on the different multi-objective problems in the context of sNTGA and cNTGA is an interesting direction for future work.

## REFERENCES

- [1] Abdoun, Otman, Jaafar Abouchabaka, and Chakir Tajani. "Analyzing the performance of mutation operators to solve the travelling salesman problem." arXiv preprint arXiv:1203.3099 (2012).
- [2] Hoffman, Karla L., Manfred Padberg, and Giovanni Rinaldi. "Traveling salesman problem." Encyclopedia of operations research and management science (2013): 1573-1578.

TABLE II  
AVERAGED VALUES OF QMS FOR SELECTED DATA INSTANCES AND ALL INVESTIGATED METHODS

Instance	Method	ED		HV [ $10^8$ ]		PFS		S		RNI	
		Avg	Std	Avg	Std	Avg	Std	Avg	Std	Avg	Std
kroAB100	NSGA-II	0.19726	0.00504	0.86063	0.00426	499.96	0.20	0.00062	0.00026	0.000050	0.000000
	NTGA	<b>0.18122</b>	<b>0.00705</b>	0.84713	0.00662	928.16	294.22	0.00154	0.00060	0.000093	0.000029
	sNTGA	0.24831	0.00999	0.88968	0.00091	2816.50	481.15	0.00041	0.00005	0.000282	0.000048
	cNTGA	0.24838	0.00880	0.88959	0.00103	<b>3062.04</b>	<b>451.13</b>	<b>0.00037</b>	<b>0.00004</b>	<b>0.000306</b>	<b>0.000045</b>
	MOEA/NSM	0.23644	0.00229	<b>0.88974</b>	<b>0.00068</b>	134.46	3.80	0.00295	0.00020	0.000013	0.000000
kroAB200	NSGA-II	0.15208	0.00209	0.86446	0.00318	499.92	0.56	0.00031	0.00014	0.000038	0.000000
	NTGA	<b>0.14295</b>	<b>0.00363</b>	0.85400	0.00418	1178.50	319.39	0.00147	0.00064	0.000091	0.000025
	sNTGA	0.24601	0.01243	0.91769	0.00078	2895.34	587.05	0.00031	0.00004	0.000223	0.000045
	cNTGA	0.23118	0.00960	0.91756	0.00072	<b>2937.00</b>	<b>512.92</b>	<b>0.00026</b>	<b>0.00004</b>	<b>0.000226</b>	<b>0.000039</b>
	MOEA/NSM	0.21721	0.00416	<b>0.91778</b>	<b>0.00064</b>	118.92	9.13	0.00305	0.00020	0.000009	0.000001
euclidAB100	NSGA-II	0.22915	0.00593	0.83400	0.00560	500.00	0.00	0.00061	0.00027	0.000050	0.000000
	NTGA	<b>0.21413</b>	<b>0.00879</b>	0.82423	0.00551	765.86	187.23	0.00143	0.00050	0.000077	0.000019
	sNTGA	0.27353	0.00808	<b>0.86446</b>	<b>0.00083</b>	2104.08	376.61	0.00043	0.00005	0.000210	0.000038
	cNTGA	0.27143	0.00706	0.86375	0.00101	<b>2238.10</b>	<b>361.25</b>	<b>0.00041</b>	<b>0.00005</b>	<b>0.000224</b>	<b>0.000036</b>
	MOEA/NSM	0.25528	0.00254	0.86419	0.00058	122.04	6.35	0.00264	0.00018	0.000012	0.000001
euclidAB300	NSGA-II	<b>0.15484</b>	<b>0.00210</b>	0.84077	0.00335	499.92	0.34	0.00031	0.00021	0.000031	0.000000
	NTGA	0.20411	0.00287	0.81625	0.00215	282.52	20.00	0.00119	0.00039	0.000018	0.000001
	sNTGA	0.24592	0.00805	0.91417	0.00128	<b>2044.48</b>	<b>344.46</b>	0.00038	0.00006	<b>0.000128</b>	<b>0.000022</b>
	cNTGA	0.24155	0.00640	0.91322	0.00108	1891.90	281.66	0.00038	0.00008	0.000118	0.000018
	MOEA/NSM	0.25118	0.00426	<b>0.91606</b>	<b>0.00065</b>	70.86	4.21	<b>0.00345</b>	<b>0.00035</b>	0.000004	0.000000
euclidAB500	NSGA-II	<b>0.13579</b>	<b>0.00335</b>	0.84135	0.00425	493.62	28.90	<b>0.00026</b>	<b>0.00024</b>	0.000022	0.000001
	NTGA	0.20428	0.00272	0.79587	0.00221	311.54	20.54	0.00086	0.00026	0.000014	0.000001
	sNTGA	0.23147	0.01084	0.91348	0.00370	<b>1236.66</b>	<b>104.87</b>	0.00060	0.00013	<b>0.000056</b>	<b>0.000005</b>
	cNTGA	0.24321	0.00796	0.91286	0.00244	1158.04	119.49	0.00062	0.00015	0.000053	0.000005
	MOEA/NSM	0.25053	0.00432	<b>0.92940</b>	<b>0.00052</b>	71.00	3.63	0.00385	0.00030	0.000003	0.000000
euclidCD100	NSGA-II	0.22927	0.00579	0.83732	0.00461	499.84	1.12	0.00058	0.00021	0.000050	0.000000
	NTGA	<b>0.22511</b>	<b>0.01555</b>	0.82967	0.00487	813.28	262.10	0.00142	0.00045	0.000081	0.000026
	sNTGA	0.27988	0.00796	<b>0.86868</b>	<b>0.00082</b>	2165.74	372.63	0.00045	0.00005	0.000217	0.000037
	cNTGA	0.27551	0.00883	0.86785	0.00103	<b>2308.40</b>	<b>347.60</b>	<b>0.00042</b>	<b>0.00006</b>	<b>0.000231</b>	<b>0.000035</b>
	MOEA/NSM	0.25717	0.00210	0.86806	0.00089	124.74	4.49	0.00275	0.00016	0.000012	0.000000
euclidCD300	NSGA-II	<b>0.15006</b>	<b>0.00167</b>	0.84324	0.00413	499.86	0.75	<b>0.00030</b>	<b>0.00028</b>	0.000031	0.000000
	NTGA	0.19801	0.00301	0.82061	0.00233	280.48	19.78	0.00102	0.00022	0.000018	0.000001
	sNTGA	0.23836	0.00792	0.91524	0.00101	<b>2004.44</b>	<b>402.02</b>	0.00037	0.00006	<b>0.000125</b>	<b>0.000025</b>
	cNTGA	0.23630	0.00694	0.91429	0.00094	1820.68	276.83	0.00039	0.00007	0.000114	0.000017
	MOEA/NSM	0.24571	0.00345	<b>0.91729</b>	<b>0.00054</b>	71.90	6.03	0.00347	0.00036	0.000004	0.000000
euclidCD500	NSGA-II	<b>0.13560</b>	<b>0.00294</b>	0.84205	0.00339	498.18	7.65	<b>0.00023</b>	<b>0.00028</b>	0.000023	0.000000
	NTGA	0.20330	0.00282	0.79741	0.00236	319.60	24.18	0.00078	0.00025	0.000015	0.000001
	sNTGA	0.23244	0.00873	0.91342	0.00319	<b>1252.72</b>	<b>117.90</b>	0.00059	0.00012	<b>0.000057</b>	<b>0.000005</b>
	cNTGA	0.24222	0.00694	0.91302	0.00246	1164.88	105.37	0.00057	0.00007	0.000053	0.000005
	MOEA/NSM	0.24629	0.00383	<b>0.92956</b>	<b>0.00031</b>	71.40	1.85	0.00405	0.00041	0.000003	0.000000
euclidEF100	NSGA-II	0.21616	0.00468	0.84541	0.00425	499.94	0.31	0.00051	0.00012	0.000050	0.000000
	NTGA	<b>0.20595</b>	<b>0.01472</b>	0.83559	0.00686	907.94	276.57	0.00118	0.00033	0.000091	0.000028
	sNTGA	0.25322	0.00778	0.87290	0.00084	2252.02	356.98	0.00043	0.00005	0.000225	0.000036
	cNTGA	0.24937	0.00871	0.87260	0.00100	<b>2380.86</b>	<b>393.59</b>	<b>0.00042</b>	<b>0.00005</b>	<b>0.000238</b>	<b>0.000039</b>
	MOEA/NSM	0.24236	0.00249	<b>0.87311</b>	<b>0.00065</b>	124.38	5.19	0.00258	0.00018	0.000012	0.000001
euclidEF300	NSGA-II	<b>0.15452</b>	<b>0.00225</b>	0.83933	0.00347	499.94	0.24	<b>0.00030</b>	<b>0.00018</b>	0.000031	0.000000
	NTGA	0.20415	0.00259	0.81662	0.00218	282.42	20.09	0.00110	0.00025	0.000018	0.000001
	sNTGA	0.24362	0.00937	0.91296	0.00091	<b>2111.34</b>	<b>462.54</b>	0.00037	0.00006	<b>0.000132</b>	<b>0.000029</b>
	cNTGA	0.23730	0.00765	0.91214	0.00080	1831.44	280.77	0.00038	0.00006	0.000114	0.000018
	MOEA/NSM	0.24726	0.00395	<b>0.91467</b>	<b>0.00059</b>	70.00	4.14	0.00327	0.00032	0.000004	0.000000
euclidEF500	NSGA-II	<b>0.13630</b>	<b>0.00316</b>	0.84093	0.00374	494.20	21.96	<b>0.00024</b>	<b>0.00025</b>	0.000022	0.000001
	NTGA	0.20351	0.00259	0.79635	0.00238	318.98	20.26	0.00083	0.00029	0.000014	0.000001
	sNTGA	0.23069	0.00885	0.91223	0.00321	<b>1207.88</b>	<b>112.44</b>	0.00056	0.00012	<b>0.000055</b>	<b>0.000005</b>
	cNTGA	0.24066	0.00608	0.91137	0.00223	1114.08	100.15	0.00057	0.00007	0.000051	0.000005
	MOEA/NSM	0.24131	0.00376	<b>0.92867</b>	<b>0.00050</b>	70.20	2.89	0.00368	0.00040	0.000003	0.000000

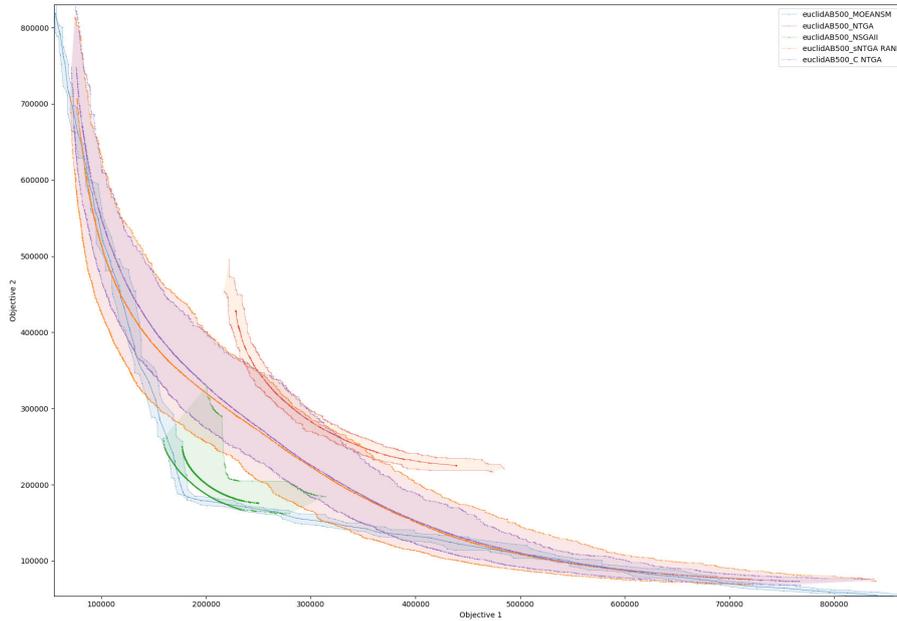


Fig. 2. Comparison of average approx. Pareto Fronts for data instance euclidAB500

TABLE III

AVERAGED RESULTS COMPARISON FOR ALL INVESTIGATED METHODS

	min ED	max HV [ $10^8$ ]	max PFS	min S	max RNI
NSGA-II	0.17191	0.84450	498.67	<b>0.00039</b>	0.00004
NTGA	0.19879	0.82125	580.84	0.00117	0.00005
sNTGA	<b>0.24759</b>	<b>0.89954</b>	<b>2008.29</b>	<b>0.00045</b>	<b>0.00016</b>
cNTGA	<b>0.24701</b>	0.89893	1991.58	<b>0.00044</b>	<b>0.00016</b>
MOEA/NSM	<b>0.24461</b>	<b>0.90441</b>	95.45	0.00325	0.00001

- [3] Cai, Xinye, et al. "An adaptive memetic framework for multi-objective combinatorial optimization problems: studies on software next release and travelling salesman problems." *Soft Comp.* 21.9 (2017): 2215-2236.
- [4] Deb, Kalyanmoy, et al. "A fast and elitist multiobjective genetic algorithm: NSGA-II." *IEEE transactions on evolutionary computation* 6.2 (2002): 182-197.
- [5] Laszczyk, Maciej, and Paweł B. Myszowski. "Improved selection in evolutionary multi-objective optimization of Multi-Skill Resource-Constrained project scheduling problem." *Information Sciences* 481 (2019): 412-431.
- [6] Laszczyk, Maciej, and Paweł B. Myszowski. "Survey of quality measures for multi-objective optimization. Construction of complementary set of multi-objective quality measures." *Swarm and Evolutionary Computation* 48 (2019): 109-133.
- [7] Miller, Donald L., and Joseph F. Pekny. "Exact solution of large asymmetric traveling salesman problems." *Science* 251.4995 (1991): 754-761.
- [8] Bolanos, R., M. Echeverry, and J. Escobar. "A multiobjective non-dominated sorting genetic algorithm (NSGA-II) for the Multiple Traveling Salesman Problem." *Decision Science Letters* 4.4 (2015): 559-568.
- [9] Elgesem, Aurora Smith, et al. "A traveling salesman problem with pickups and deliveries and stochastic travel times: An application from chemical shipping." *European Journal of Operational Research* 269.3

(2018): 844-859.

- [10] Braekers, Kris, Katrien Ramaekers, and Inneke Van Nieuwenhuysse. "The vehicle routing problem: State of the art classification and review." *Computers & Industrial Engineering* 99 (2016): 300-313.
- [11] Maity, Samir, Arindam Roy, and Manoranjan Maiti. "An imprecise multi-objective genetic algorithm for uncertain constrained multi-objective solid travelling salesman problem." *Expert Systems With Applications* 46 (2016): 196-223.
- [12] Moraes, Deyvid Heric, et al. "A novel multi-objective evolutionary algorithm based on subpopulations for the bi-objective traveling salesman problem." *Soft Computing* (2018): 1-12.
- [13] Seada, Haitham, Mohamed Abouhawwash, and Kalyanmoy Deb. "Multiphase Balance of Diversity and Convergence in Multiobjective Optimization." *IEEE Transactions on Evolutionary Computation* 23.3 (2018): 503-513.
- [14] Ariyasingha, I. D. I. D., and T. G. I. Fernando. "Performance analysis of the multi-objective ant colony optimization algorithms for the traveling salesman problem." *Swarm and Evolutionary Comp.* 23 (2015): 11-26.
- [15] Ke, Liangjun, Qingfu Zhang, and Roberto Battiti. "MOEA/D-ACO: A multiobjective evolutionary algorithm using decomposition and ant-colony." *IEEE transactions on cybernetics* 43.6 (2013): 1845-1859.
- [16] Ke, Liangjun, Qingfu Zhang, and Roberto Battiti. "Hybridization of decomposition and local search for multiobjective optimization." *IEEE transactions on cybernetics* 44.10 (2014): 1808-1820.
- [17] Chen, Xinye, et al. "Ant colony optimization based memetic algorithm to solve bi-objective multiple traveling salesmen problem for multi-robot systems." *IEEE Access* 6 (2018): 21745-21757.
- [18] Cornu, Marek, Tristan Cazenave, and Daniel Vanderpooten. "Perturbed decomposition algorithm applied to the multi-objective traveling salesman problem." *Computers & Operations Research* 79 (2017): 314-330.
- [19] Lust, Thibaut, and Jacques Teghem. "The multiobjective traveling salesman problem: a survey and a new approach." *Advances in Multi-Objective Nature Inspired Computing*. Springer, Berlin, Heidelberg, 2010. 119-141.
- [20] Qamar, Nosheen, Nadeem Akhtar, and Irfan Younas. "Comparative Analysis of Evolutionary Algorithms for Multi-Objective Travelling

TABLE IV  
AVERAGED VALUES OF PURITY FOR SELECTED INVESTIGATED METHODS

Instance	sNTGA vs cNTGA		cNTGA vs sNTGA		sNTGA vs MOEA/NSM		MOEA/NSM vs sNTGA		cNTGA vs MOEA/NSM		MOEA/NSM vs cNTGA	
	Avg	Std	Avg	Std	Avg	Std	Avg	Std	Avg	Std	Avg	Std
	euclidAB100	0.397	0.187	0.652	0.187	0.450	0.173	0.687	0.166	0.571	0.141	0.595
euclidAB300	0.275	0.181	0.656	0.210	0.216	0.076	0.728	0.066	0.294	0.102	<b>0.714</b>	0.082
euclidAB500	0.391	0.363	0.607	0.356	0.716	0.216	0.369	0.210	<b>0.854</b>	0.140	0.240	0.138
euclidCD100	0.384	0.204	0.642	0.202	0.431	0.156	0.694	0.141	0.534	0.187	0.616	0.163
euclidCD300	0.243	0.152	0.695	0.180	0.201	0.068	0.748	0.064	0.305	0.094	<b>0.714</b>	0.075
euclidCD500	0.462	0.327	0.540	0.322	0.750	0.187	0.336	0.172	<b>0.863</b>	0.124	0.233	0.130
euclidEF100	0.420	0.216	0.623	0.195	0.485	0.183	0.647	0.147	0.563	0.187	0.584	0.161
euclidEF300	0.239	0.196	0.684	0.231	0.187	0.052	0.751	0.059	0.293	0.083	<b>0.720</b>	0.076
euclidEF500	0.393	0.298	0.605	0.293	0.788	0.188	0.293	0.190	<b>0.922</b>	0.085	0.154	0.093
kroAB100	0.448	0.210	0.597	0.210	0.453	0.177	0.669	0.146	0.527	0.194	0.594	0.185
kroAB200	0.447	0.251	0.513	0.251	0.230	0.123	0.655	0.138	0.251	0.116	<b>0.771</b>	0.122
Average	0.372	0.235	<b>0.620</b>	0.240	0.446	0.146	<b>0.598</b>	0.136	<b>0.543</b>	0.132	<b>0.539</b>	0.125

Salesman Problem." INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS 9.2 (2018): 371-379.

- [21] Moraes, Deyvid Heric, et al. "A novel multi-objective evolutionary algorithm based on subpopulations for the bi-objective traveling salesman problem." *Soft Computing* (2018): 1-12.
- [22] Hassin, Refael, and Shlomi Rubinstein. "Better approximations for max TSP." *Information Processing Letters* 75.4 (2000): 181-186.
- [23] Abdoun, Otman, and Jaafar Abouchabaka. "A comparative study of adaptive crossover operators for genetic algorithms to resolve the traveling salesman problem." *arXiv preprint arXiv:1203.3097* (2012).
- [24] Varun Kumar, S. G., and R. Panneerselvam. "A study of crossover operators for genetic algorithms to solve VRP and its variants and new sinusoidal motion crossover operator." *Int. J. Comput. Intell. Res* 13.7 (2017): 1717-1733.
- [25] Stehling, Thiago Muniz, and Sergio Ricardo de Souza. "A Comparison of Crossover Operators Applied to the Vehicle Routing Problem with Time Window." *2017 Brazilian Conference on Intelligent Systems (BRACIS)*. IEEE, 2017.
- [26] Lopez-Ibanez M., Paquete L., and Stutzle T. "Exploratory Analysis of Stochastic Local Search Algorithms in Biobjective Optimization", *Experimental Methods for the Analysis of Opt. Alg.* (2010): 209-222.
- [27] Whitley, L. Darrell, Timothy Starkweather, and D'Ann Fuquay. "Scheduling problems and traveling salesmen: The genetic edge recombination operator." *ICGA*. Vol. 89. 1989.

# Development of a Flexible Mizar Tokenizer and Parser for Information Retrieval System

Kazuhisa Nakasho  
 Yamaguchi University  
 in Yamaguchi

2-16-1, Tokiwa-dai, Ube City, Yamaguchi, Japan  
 Email: nakasho@yamaguchi-u.ac.jp

**Abstract**—In this paper, we explain the development of a new Mizar tokenizer and parser program as a component of a search system that works on the Mizar Mathematical Library. The existing Mizar tokenizer and parser can handle only an article as a whole written in the Mizar language, however, the newly developed program can deal with a snippet of a Mizar article.

In particular, since it is possible to handle a snippet of an article without specifying a vocabulary section of an environment part, it is expected that user input efforts will be greatly reduced.

## I. MOTIVATION

THE AUTHOR is developing a new information retrieval system that works on the Mizar Mathematical Library (MML) [1]. In this paper, we explain a developed tokenizer and parser program of the Mizar language as a component of our search system.

### A. MML Query

Currently, MML Query [2] developed by Grzegorz Bancerek in 2001 is widely used as an MML theorem search system. MML Query is the forerunner of the search systems for formalized mathematical libraries. Even today, it is the only active system that can search comprehensively large formalized mathematical libraries [3]. MML Query realizes pattern matching according to the grammatical structure of the Mizar language with its own language to specify a search object. This feature allows the users to input search patterns that have more expressive power than that of regular expressions. However, the users have to spend a considerable amount of time to learn the grammar of its own search language. Furthermore, since a mathematical theorem can be transformed into an infinite number of patterns by equivalent rewriting, it often causes retrieval omission in pattern matching. As mentioned above, MML Query has succeeded in reducing laborious retrieval work in the MML, however, there is still rooms for improvement.

### B. Developing search system

In order to learn from the drawbacks of MML Query, the newly developing search system extracts features of the input data and compares them with that of theorems and definitions registered in the MML. For the feature comparison, we use an algorithm designed to output a logical distance between two expressions. In addition, the search history will be collected

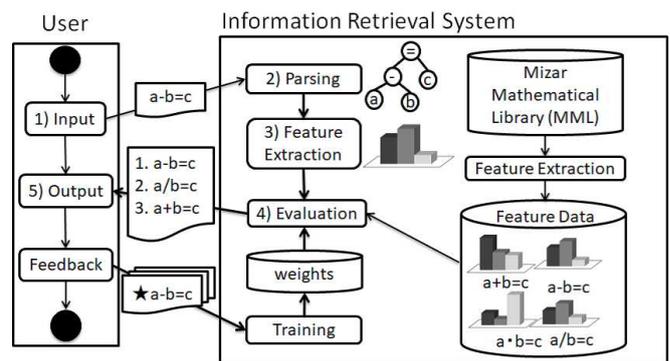


Fig. 1. Diagram of our information retrieval system

and reused as corpus data for machine learning so that the distance calculation algorithm will be tuned to fit user trend.

The flow of our search system is as follows:

- 1) A user inputs a search target such as a theorem in the Mizar language.
- 2) The search system parses the input data.
- 3) The search system calculates the features of the input data such as a number of occurrences and positional relationship of symbols and variables by analyzing the syntax tree.
- 4) The search system compares the above features with that of theorems and definitions registered in the MML, and ranks and displays the matching rates.
- 5) The user of the search system browses the search result displayed on the system.

Fig. 1 shows a diagram of our search system.

### C. Necessity of a new tokenizer and parser

The developed tokenizer and parser program correspond to process 2) in Fig. 1. The existing parser used in the proof verification program of the Mizar system first reads an article and converts it into the Weakly Strict Mizar (WS-Mizar) language [4], [5]. In the WS-Mizar language, all terms are fully parenthesized, therefore, there is no ambiguity in operator precedence. After this process, the article in the WS-Mizar language is converted into XML intermediate representation by a parser program generated with Bison. The existing parser

program is supposed to handle full text of a Mizar article, therefore, it cannot process a snippet like a theorem, which is expected as the main input of our search system. That is the reason why we needed to develop a new tokenizer and parser program for the Mizar language.

## II. REQUIREMENTS OF NEW TOKENIZER AND PARSER

The Mizar language consists of a context sensitive grammar, and a set of valid symbols are determined according to Mizar articles enumerated in a vocabulary section of an environment part. However, in the construction of a Mizar article, it is said that the most difficult process is to create an environment part correctly. Therefore, it is not practical to enforce a search system user to input an environment part for every search. In this project, we aimed at constructing a tokenizer and parser program that works practically without an environment part. However, since the omission of an environment part may cause unexpected syntax errors, our search system needs to provide interfaces that enable the users to grasp and correct any syntax errors easily.

### A. Tokenizer

As mentioned earlier, in the Mizar language, valid symbols are determined according to the Mizar articles enumerated in a vocabulary section. It means that a vocabulary section has an ability to determine word boundaries in lexical analysis. When a vocabulary section is omitted, our tokenizer extracts tokens according to the longest match rule on the assumption that every symbol registered in the MML is valid. As a result, a token that is not an originally valid symbol might be mistakenly recognized as a symbol. However, we succeeded in reducing token recognition errors by implementing special interpretation rules to recognize a token placed immediately after a certain keyword such as *let* or *reserve* as a variable.

### B. Parser

Bison, which is used as a parser generator for the existing Mizar parser, adopts LALR parsing known as one of the most practical bottom-up parsing algorithms. Generally, bottom-up parser generators produce more efficient and smaller programs than top-down parser generators. However, since bottom-up parsers have difficulty in constructing a rough tree structure in the middle of parsing process, they sometimes tend to output meaningless error messages when grammatically incorrect input is given. The existing Mizar proof verification system also tends to output grammatical error messages that are difficult for beginners to understand. Based on these reasons and recent performance improvement of computer hardware, there have been increasing cases where top-down parsers such as LL parser and packrat parser are used in recent years. We also adopted ANTLR, which is based on Adaptive LL (\*) parsing and known as one of the most powerful top-down parser generators, because incomplete input will often be given to our search system. ANTLR supports many output languages such as Java, C++, Python2, Python3, Go, Swift, JavaScript and C#. Whenever we need to develop Mizar tools that work

on Web browser or modern editors such as Atom or Visual Studio Code, it can generate a parser written in JavaScript immediately.

## III. PROGRAM SPECIFICATION

This section explains input, output and the flow of our tokenizer and parser program. Fig. 2 shows the flow chart of our program.

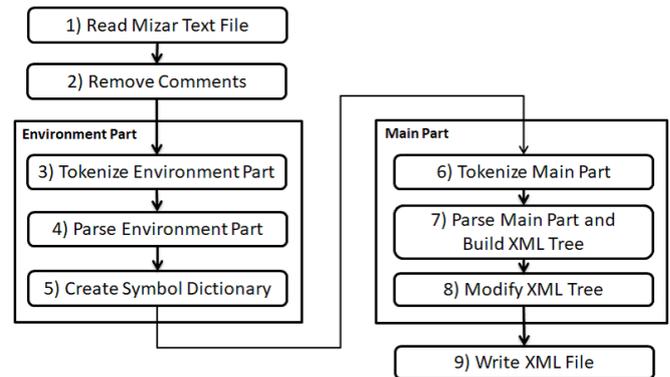


Fig. 2. Flow of our tokenizer and parser program

### A. Input and output

Our program accepts not only full text of a Mizar article but also various types of blocks such as theorem, definition, registration, notation, and scheme as input data. Our program outputs a parsing result in XML format as well as the existing Mizar parser program. The output XML faithfully reproduces the structure of the official BNF grammar provided at [mizar.org](http://mizar.org). Applying the official grammar rules to output XML will promote the secondary use of our tokenizer and parser program.

### B. Tokenizer specification

In lexical analysis, when a vocabulary section is not specified, it is assumed that all the symbols defined in the MML are valid.

All the symbols in the MML are recorded in *mml.vct* attached to Mizar distribution binaries. In our tokenizer program, a symbol dictionary is built by extracting symbol information from *mml.vct* in a pre-processing step.

Our tokenizer program first removes comments. Next, it reads tokens such as symbols, keywords, numbers, identifiers, etc. from the left hand side according to the longest match rule. If the token matches a symbol registered in the symbol dictionary, our tokenizer program appends prefix "`__` «symbol type» «symbol priority» `_`" to the token. Owing to the prefix, the following parser program is able to distinguish symbol types in the parsing process. Table I shows the correspondence between symbol types and their meanings in the Mizar language.

When a token is cut out according to the longest match rule on the assumption that all the MML symbols are valid, there

TABLE I  
CORRESPONDENCE BETWEEN SYMBOL TYPES AND THEIR MEANINGS

Symbol type	Meaning
R	Predicate
O	Functor
M	Mode
G	Structure
U	Selector
V	Attribute
K	Left Functor Bracket
L	Right Functor Bracket

is a risk that a variable is misinterpreted as a symbol. For this reason, when a token comes immediately after a certain keyword such as *let* or *reserve*, our tokenizer regards the token as a variable identifier, and its symbol validity is temporarily turned off within the scope of the variable. Our program writes out token-separated text at the end of the process.

### C. Parser specification

The official syntax of the Mizar language is written in BNF. We transformed the BNF syntax definition into ANTLR grammar form, then passed it to ANTLR parser generator. Normally, LL parsers require left recursion removal, although ANTLR automatically resolves direct left recursions. For this reason, we only needed to remove indirect left recursions. There are only two indirect left recursion in the Mizar official syntax definition. We repaired them and transformed the grammar rules from BNF to ANTLR grammar form.

The Mizar language has a feature that allows users to define prioritized infix operators (functors). While this feature has given a significant advantage of the readability of the Mizar language, it has also made lexical and syntactic analysis more difficult. Historically, this grammatical complexity has often become a bottleneck in the development of support tools for the Mizar system [6]. The existing Mizar system converts the Mizar language into the WS-Mizar language, thus all terms are parenthesized. Thanks to this process, the existing parser can avoid ambiguity in associativity and precedence of infix operators. In our program, its syntax tree structure is re-edited according to infix operator priorities in post-process. This strategy is also used in a parser of Standard ML [7]. At the end of parsing process, our parser removes prefixes attached by a tokenizer to symbols, then outputs an XML file.

### D. Choice of programming languages

We selected C++ for our parser implementation language because the parsing process takes most of the execution time and requires a high performance implementation. On the other hand, we also chose Python3 for other processes to realize smooth linkage with other programs and increase the productivity of the implementation. As for the parser, when we tried both C++ and Python3 as ANTLR output languages, we confirmed that C++ is about 10 times faster than Python3. In the C++ version, the parsing process occupies about 50 to 70 percent of the whole execution time. To bridge the gap

between Python3 and C++, we used the C++ extension feature of Python3 so that the data exchange is performed on memory.

## IV. EVALUATION

The source code of our program is published and managed on GitHub under the MIT license<sup>1</sup>.

### A. Functionality

Most of our program is written in Python3 and is composed of highly extensible modules. Furthermore, since our program faithfully outputs an XML file that follows the official grammar rules written in BNF, it is easy to reuse its source code for development of other support tools of the Mizar system. Currently, although the platform on which our program runs is limited to UNIX, we also plan to support Windows and Mac OS in the future.

### B. Performance

Table II shows a number of words and file size of each Mizar file used for a performance test. *jordan:95* is input data for a test case of our search system so that this file consists of a single theorem and a vocabulary section is not included. The file size of *ring\_1* is standard and that of *jgraph\_4* is the largest in the MML, respectively.

TABLE II  
SPECIFICATION OF MIZAR ARTICLES

	number of words	size
<i>jordan:95</i>	168	0.575 kB
<i>ring_1</i>	11558	37.6 kB
<i>jgraph_4</i>	185895	492 kB

Table III shows the specification of the test environment used in the performance test.

TABLE III  
TEST ENVIRONMENT

Item	Specification
CPU	Intel®Core™i7-7500U @ 2.70 GHz
Memory	16 GB
OS	Ubuntu 16.04 LTS
Compiler	GCC version 7

Table IV shows the execution time of each step of our program. Each item in the list corresponds to the labelled process shown in Fig. 2. From this table, it is confirmed that most of the execution time is occupied by process 7), that is, occupied by an ANTLR generated parser. According to the measurement results, in the case of an article of about 10,000 words like *ring\_1*, the time consumption is less than one second, which means it is enough to be used in practical applications. However, when it comes to an article with more than 100,000 words like *jgraph\_4*, parsing time exceeds 10 seconds. Currently, we suppose the application of our program is limited to the analysis of small input data

TABLE IV  
TIME CONSUMPTION OF EACH PROCESS

	jordan:95	ring_1	jordan_1
1) Read Mizar Text File	0.0029 s	0.0004 s	0.0102 s
2) Remove Comments	0.0000 s	0.0007 s	0.0039 s
3) Tokenize Environment Part	0.0000 s	0.0019 s	0.0026 s
4) Parse Environment Part	0.0000 s	0.0030 s	0.0017 s
5) Create Symbol Dictionary	0.0196 s	0.0061 s	0.0077 s
6) Tokenize Environment Part	0.0015 s	0.1054 s	2.2488 s
7) Parse Main Part & Build XML	0.0235 s	0.4444 s	15.1322 s
8) Modify XML Tree	0.0013 s	0.0731 s	1.3009 s
9) Write XML File	0.0022 s	0.0275 s	0.4751 s
Total	0.0511 s	0.6626 s	19.1831 s

like *jordan:95*. Hence, we conclude our program already has enough performance for the application.

Table V shows the comparison result of performance measurement between our program and the existing Mizar parser. Even though this performance comparison is unfair because the existing Mizar parser has additional features such as indexing variables, it is enough to check approximate relative performance of these two programs. This comparison made it clear that our program tends to be slower than the existing parser as the input file size becomes larger. This tendency mainly comes from the difference of parser algorithms between the conventional bottom-up parsing and top-down parsing. The official grammar rules of the Mizar language include a significant number of left recursions that cause backtracking in the process of top-down LL(\*) parsing and performance deterioration. This performance deterioration is a well known issue that occurs when ANTLR generates parsers for languages with complex grammar rules.

TABLE V  
PERFORMANCE COMPARISON BETWEEN CURRENT AND NEW VERSIONS

	current version	new version
ring_1	3.013 s	0.662 s
jgraph_4	3.610 s	19.183 s

## V. REMAINING CHALLENGES

### A. Display of parsing results

Since token interpretation of the Mizar language depends on the entries in a vocabulary section of an environment part, there is a possibility that the program produces incorrect results against user intention when it is applied to input data without a vocabulary section. Therefore, when parsing a snippet by our program, it is necessary to provide a graphical user interface (GUI) that allows users to check the parsed result visually. In the development of our search system, we are planning to build up a component that converts a parsing result into an HTML document with highlights of syntax errors, hyperlinks to symbol definitions, and so on.

<sup>1</sup><https://github.com/mimosa-project/emparser>

### B. Type checking

The Mizar language allows symbol overloading and mode inheritance as well as Java and C++ languages. Therefore, type checking or type inference must be performed in the semantic analysis. Improving the precision of the semantic analysis is expected to greatly contribute to improve on the accuracy of our search system. There is a preceding research on type inference without an environment part by Cezary Kaliszyk et al. [8].

### C. Performance improvement

We are planning to improve the performance for the case where our program is applied to other than our search system in the future. According to the performance measurements, it is supposed that effective remedies are to change the parser algorithm to bottom-up parsing or to optimize a grammar file passed to ANTLR. However, the replacement to a bottom-up parsing makes error handling more difficult, and optimization of the ANTLR grammar file has a disadvantage of impairing readability of the syntax rules. Another improvement plan is to rewrite program components written in Python, such as tokenizer, by using C++ extensions.

### ACKNOWLEDGMENT

I would like to express my gratitude to Adam Naumowicz, Artur Korniłowicz and Radosław Piliszek, who explained the specification of the existing Mizar tokenizer and parser programs and provided a part of their source code.

### REFERENCES

- [1] G. Bancerek, C. Byliński, A. Grabowski, A. Korniłowicz, R. Matuszewski, A. Naumowicz, and K. Pał, "The role of the Mizar Mathematical Library for interactive proof development in Mizar," *Journal of Automated Reasoning*, vol. 61, no. 1, pp. 9–32, Jun 2018. [Online]. Available: <https://doi.org/10.1007/s10817-017-9440-6>
- [2] G. Bancerek, "Information retrieval and rendering with MML query," in *Mathematical Knowledge Management*, J. M. Borwein and W. M. Farmer, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 266–279. [Online]. Available: [https://doi.org/10.1007/11812289\\_21](https://doi.org/10.1007/11812289_21)
- [3] F. Guidi and C. Sacerdoti Coen, "A survey on retrieval of mathematical knowledge," *Mathematics in Computer Science*, vol. 10, no. 4, pp. 409–427, Dec 2016. [Online]. Available: <https://doi.org/10.1007/s11786-016-0274-0>
- [4] C. Bylinski and J. Alama, "New developments in parsing Mizar," in *Intelligent Computer Mathematics*, J. Jeuring, J. A. Campbell, J. Carette, G. Dos Reis, P. Sojka, M. Wenzel, and V. Sorge, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 427–431. [Online]. Available: [https://doi.org/10.1007/978-3-642-31374-5\\_30](https://doi.org/10.1007/978-3-642-31374-5_30)
- [5] A. Naumowicz and R. Piliszek, "Accessing the Mizar library with a weakly strict Mizar parser," in *Intelligent Computer Mathematics*, M. Kohlhase, M. Johansson, B. Miller, L. de Moura, and F. Tompa, Eds. Cham: Springer International Publishing, 2016, pp. 77–82. [Online]. Available: [https://doi.org/10.1007/978-3-319-42547-4\\_6](https://doi.org/10.1007/978-3-319-42547-4_6)
- [6] P. Cairns and J. Gow, "Integrating searching and authoring in Mizar," *Journal of Automated Reasoning*, vol. 39, no. 2, pp. 141–160, Aug 2007. [Online]. Available: <https://doi.org/10.1007/s10817-007-9073-2>
- [7] A. W. Appel and D. B. MacQueen, "Standard ML of New Jersey," in *International Symposium on Programming Language Implementation and Logic Programming*. Springer, 1991, pp. 1–13. [Online]. Available: [https://doi.org/10.1007/3-540-54444-5\\_83](https://doi.org/10.1007/3-540-54444-5_83)
- [8] C. Kaliszyk, J. Urban, and J. Vyskočil, "Learning to parse on aligned corpora (rough diamond)," in *International Conference on Interactive Theorem Proving*. Springer, 2015, pp. 227–233. [Online]. Available: [https://doi.org/10.1007/978-3-319-22102-1\\_15](https://doi.org/10.1007/978-3-319-22102-1_15)

# British Sign Language Recognition In The Wild Based On Multi-Class SVM

M. Quinn and J.I. Olszewska

School of Computing and Engineering  
 University of the West of Scotland, United Kingdom  
 Email: joanna.olszewska@ieee.org

**Abstract**—Developing assistive, cost-effective, non-invasive technologies to aid communication of people with hearing impairments is of prime importance in our society, in order to widen accessibility and inclusiveness. For this purpose, we have developed an intelligent vision system embedded on a smartphone and deployed in the wild. In particular, it integrates both computer vision methods involving Histogram of Oriented Gradients (HOG) and machine learning techniques such as multi-class Support Vector Machine (SVM) to detect and recognize British Visual Language (BSL) signs automatically. Our system was successfully tested on a real-world dataset containing 13,066 samples and shown an accuracy of over 99% with an average processing time of 170ms, thus appropriate for real-time visual signing.

## I. INTRODUCTION

THERE are 11 million people with hearing loss across the United Kingdom (UK), i.e. 1 in 6 people, with around 900,000 of these persons having profound hearing loss [1]. Users of British Sign Language (BSL) number in the 150,000 range, with more than half of them using BSL as their first language [2]. By stark contrast, there are far less registered BSL interpreters, with 1540 being publicly available on the National Registers of Communication Professionals working with Deaf and Deafblind People (NRCPP) [3].

BSL consists of 26 hand-shapes; one being correlated to each letter of the alphabet, as illustrated in Fig. 1. Each letter is formed using two hands except for the letter ‘C’, using only one hand [4].

With the current expansion of Artificial Intelligence (AI) in daily applications [5], intelligent systems can play an important role for sign language recognition (SLR).

However, despite a number of technologies developed for the automated, visual recognition of gestures within the field of Human Computer Interaction (HCI) [6], [7], only very few studies have tackled with automated BSL translation [8].

In HCI, most of the gesture recognition systems usually require special hardware equipments, such as depth camera [9] or gloves [10], which are usually not available outside a laboratory and/or have limited utility in the wild.

On the other hand, SLR systems integrating machine learning techniques such as genetic algorithms (GA) [11] or convolutional neural networks (CNN) [12] have been mainly focused on the American Sign Language (ASL) [13], which uses static single-hand poses (as opposed to BSL which uses two-handed

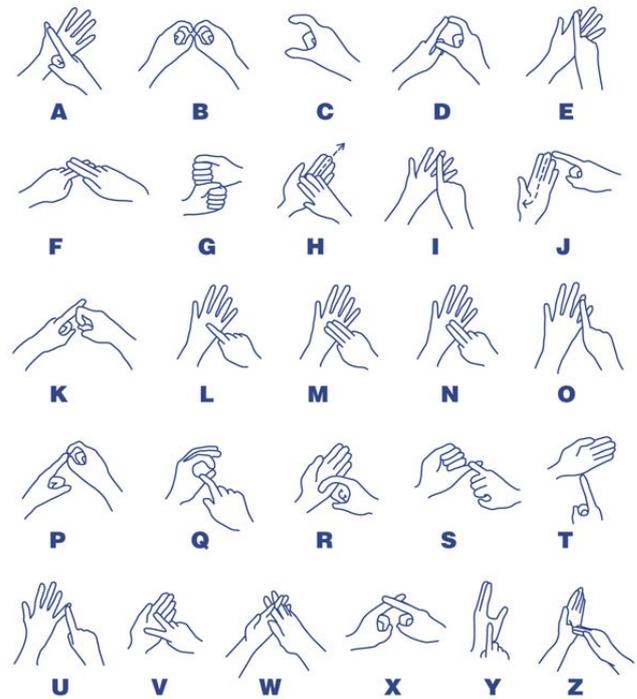


Fig. 1. Schematic overview of the British Visual Language (BSL) alphabet right-handed fingerspelling [2].

ones) to spell individual letters. Most of the available datasets are also only dedicated to ASL. It is worth noting that ASL and BSL languages have little crossover in terms of their actual constituent phonemes and are mutually incomprehensible to one another.

Hence, compared to ASL, BSL has a smaller body of research dedicated to visual recognition and as such, little online BSL datasets exist. Thus, methods based on deep learning [12], which requires hundreds of thousands of training data samples, are not directly available for BSL.

In this work, we propose the development of an accessible, intelligent vision system for real-time, automated BSL recognition in the wild. This assistive technology is inbuilt as a smartphone application, using computer-vision algorithms to process the images captured in real-time by the smartphone

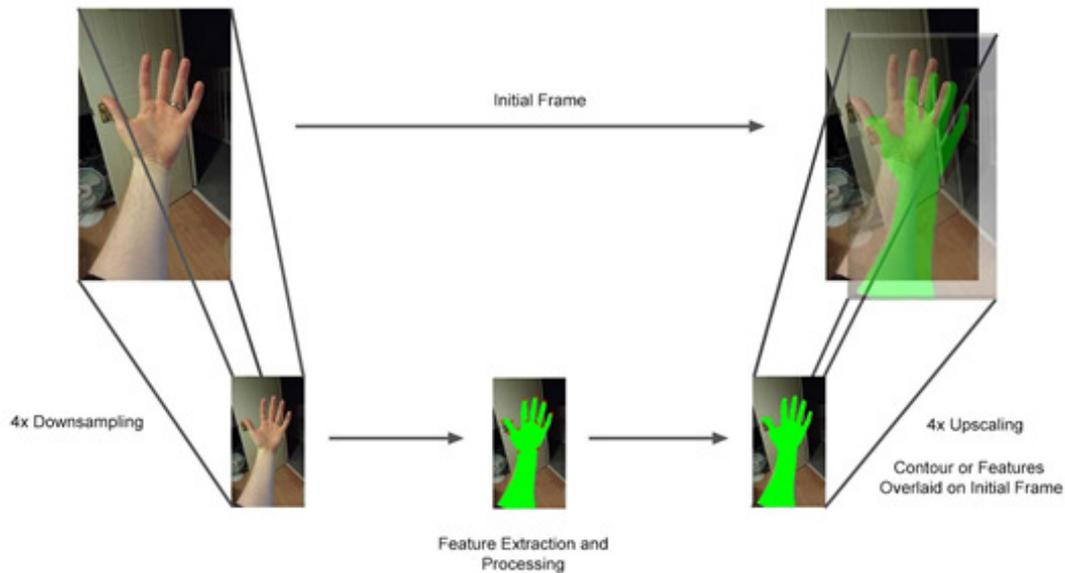


Fig. 2. Overview of our sign detection process.

camera and translating the detected hand pose into a letter using the machine learning technique called Support Vector Machine (SVM).

This intelligent vision system uses a regular optic camera such as a smartphone camera working with RGB flat image data, not requiring any type of calibration or invasive sensors.

On the other hand, our system processes images at an average rate of  $170ms$  per image, addressing successfully the real-time constraint. Indeed, the visual signing rate is of 2.3-2.5 signs per second [14], enforcing real-time SLR systems to process each sign at  $2fps$  or in less than  $400ms$  per image.

The study developed in this work aims to enrich the lives of a wide range of people with hearing and/or speech impairments and their respective circles.

The strict use of free and open source technologies in this project along with the use of cheap, portable hardware such as a smartphone implies that the resultant prototype could be highly accessible to a large number of users.

The original contribution of our work is the study of the automated BSL recognition process and includes the creation of a BSL large-scale dataset of c. 13k samples as well as the design, development, and deployment of a new intelligent vision system for automated BSL recognition in real-world environment.

The paper is structured as follows. In Section II, we present our detection and recognition system for BSL, while in Section III we report and discuss the carried out experiments which results show the developed SLR system has excellent performance on real-world large-scale datasets, both in terms of accuracy and computational efficiency. Conclusions are drawn up in Section IV.

## II. OUR METHOD

The developed intelligent vision system embeds a two-step computational process. The first step involves computer-vision algorithms processing the input image (see Fig. 2) and results in the visual sign detection, as described in Section II-A. The second step consists in machine-learning algorithms using Support Vector Machine (SVM) to recognize the detected visual sign, as explained in Section II-B.

### A. Sign Detection

Let us consider a colour RGB image or video frame  $I(x, y)$  where  $M \times N$  is its size, with  $M$ , its width, and  $N$ , its height, recorded live with the smartphone using the *OpenCV Camera Listener* triggered by our application.

In the first phase, the intelligent vision system running on the smartphone applies to the image  $I(x, y)$  a series of mathematical operations as follows.

Firstly, the RGB image is transformed to the HSV colour space [15]. Secondly, the image is resized and down-sampled based on the Gaussian Pyramid as depicted in Fig. 2. Then, the image is segmented by thresholding [16], and the mask of the hand(s) is extracted by applying mathematical morphologic operations such as eroding and dilation [17]. Next, the Histogram of Oriented Gradients (HOG) [18] is computed as shown in Fig. 3. HOG assembles a histogram of prevailing, aggregated gradients throughout predefined blocks of an image. Because HOG produces such histogram, the number of features per vector is the same every time, given the input image  $I(x, y)$  of a static size.

It is worth noting that through multiple pre-processing and cropping layers, a resultant image of a predictable size, i.e.  $M \times N$ , is given to the HOG detector feature extraction layer.

Post-processing of the image encompasses any upscaling and resizing that may need to be done to render the correct

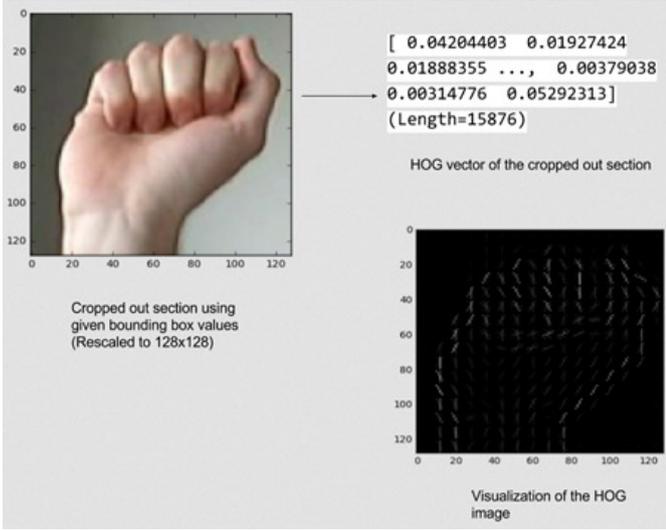


Fig. 3. Histogram of Oriented Gradients (HOG) visualisation.

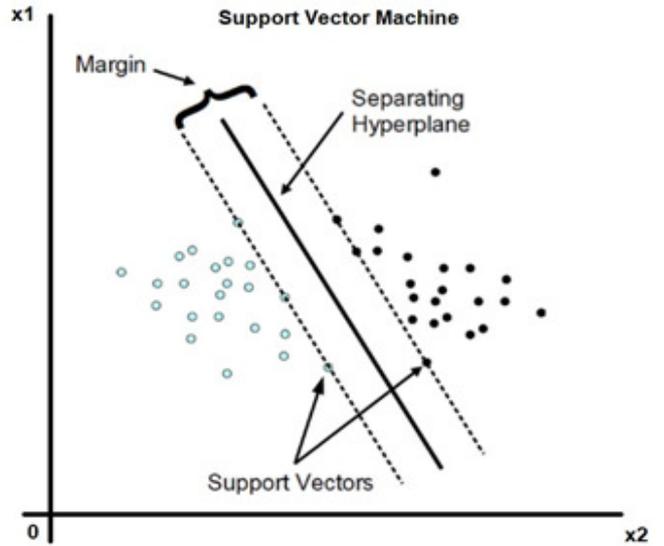


Fig. 4. Support Vector Machine (SVM) hyperplane visualisation.

information. Hence, if contour overlays are required to indicate the detected hand shape (as displayed in Fig. 8), then upscaling the previously down-sampled materials is done in this stage of the process.

*B. Sign Recognition*

Once data has been processed into a set of frames containing only pertinent gesture data as described in Section II-A, these are analysed and fed into a model as explained in this Section II-B, in order to be classified in one of the 26 classes of the BSL alphabet.

In this work, the adopted classifier is a Support Vector Machine (SVM), since SVM is an efficient implementation of a supervised machine learning approach for classification and decision making [19], while requiring only a small sample of training data [20]. SVM method is described in the subsections, as follows.

1) *SVM Hyperplane*: At its core, the SVM method attempts to find an ideal, separating hyperplane  $H_0$  (such as schematized in Fig. 4) to divide a dataset of  $n$  points into separate classes  $y_i$  (e.g. class  $y_i = +1$  and class  $y_i = -1$ ), as follows:

$$H_0 \equiv \mathbf{w}^T \mathbf{x}_i + b = 0, \tag{1}$$

with  $\mathbf{x}_i = (x_1, x_2, \dots, x_n)$ , the input vector,  $\mathbf{w} = (a, -1)$ , the weight factor,  $b$ , the bias, and  $a$  such as  $ax_1x_2 + b = 0$ ; this equation being derived from two-dimensional vectors, but in fact, works for any number  $p$  of dimensions.

The Support Vectors themselves are the data points closest to this plane of division and are therefore critical to segregating classes. Depending on the class (i.e. class +1 or class -1) they are part of, they belong either to  $H_1$  or  $H_2$ , defined as:

$$H_1 \equiv \mathbf{w}^T \mathbf{x}_i + b = 1, \tag{2}$$

$$H_2 \equiv \mathbf{w}^T \mathbf{x}_i + b = -1, \tag{3}$$

with the (hard) margin defined as  $D = H_1 - H_2$  (see Fig. 4); the hyperplane  $H_0$  being the median in between  $H_1$  and  $H_2$ .

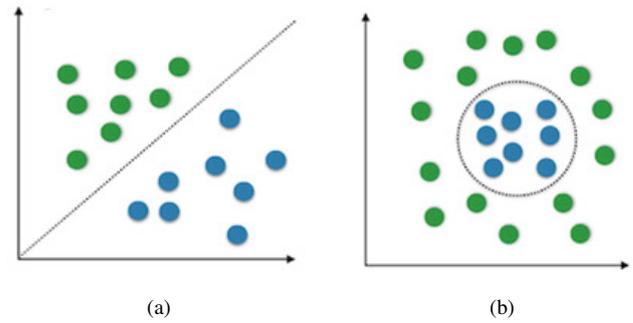


Fig. 5. SVM class separation using: (a) a linear kernel; (b) a radial basis function (RBF) kernel.



Fig. 6. Min-max normalisation performed on feature data before being fed into the SVM layer. It shows new maximum value of 1 on the lower row.

Consequently, each feature vector  $\mathbf{x}_i$  is classified as follows:

$$\text{class } y_i = +1, \text{ if } \mathbf{w}\mathbf{x}_i + b \geq 1, \tag{4}$$

$$\text{class } y_i = -1, \text{ if } \mathbf{w}\mathbf{x}_i + b \leq -1. \tag{5}$$

For SLR, this means finding a hyperplane (Eq. 1) between data from an actual gesture (i.e. class  $y_i = +1$ ) and interstitial hand movements which are not categorised (i.e. class  $y_i = -1$ ), leading to a *one-versus-all* approach, and then using the hyperplane to make predictions as per Eqs. 4-5.

2) *SVM Kernels and Hyperparameters*: The SVM is an integral part of the application, and different possible kernels could be implemented in order to gain linear separation in the data space, in case data is linearly separable (linear SVM), or otherwise (non-linear SVM) in a higher dimensional space. In particular, Linear and Radial Basis Function (RBF) kernels were used to test the application as reported

in Section III. As implied by the name, the linear kernel ( $K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j)$ ) attempt to draw a line between class data (see Fig. 5 (a)), whereas the radial kernel ( $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/2\sigma^2)$ ) tries to fit a curved or radial shape between classes (see Fig. 5 (b)). These are very different kernels, but can often produce similar results in practice.

In the finalised prototype, the RBF kernel has been adopted based on the resultant test data accuracy (see Table 1). After being trained offline with 20 samples per class, the SVM is run live within the smartphone application. The maximum iterations before terminating the SVM is set to 100. The parameters have been chosen by cross-validation on the training set.

3) *Normalisation*: Min-max normalisation has been performed on feature data before being fed into the SVM layer. Normalisation is important, since it flattens the data to an appropriate scale, in this case from 0 to 1.

The formula given for min-max normalisation is that a data point has the set's minimum subtracted from it, and then this value is divided by the set's maximum minus the set's minimum. Accordingly, a value  $A$  for  $B$  is then derived as:

$$A = \frac{B - \min(B)}{\max(B) - \min(B)}. \quad (6)$$

Figure 6 shows a small example of a row of matrix data being set to new normalised values by applying Eq. 6. Within this example, the far right non-zero, previously maximal value is set to the new scale maximal value which is 1. The other values in the matrix are then scaled according to their relation to this new maximal value.

### III. EXPERIMENTS AND DISCUSSION

Our intelligent vision system was validated by running a series of experiments and assessing them quantitatively as reported below.

The effectiveness of the prototype has been measured using the following metrics [5]:

$$\textit{precision} (P) = \frac{TP}{TP + FP}, \quad (7)$$

$$\textit{recall} (R) = \frac{TP}{TP + FN}, \quad (8)$$

$$\textit{specificity} (S) = \frac{TN}{TN + FP}, \quad (9)$$

$$\textit{accuracy} (ACC) = \frac{TP + TN}{TP + TN + FP + FN}, \quad (10)$$

where  $TP$  is the True positive rate,  $FP$  is the False Positive rate,  $FN$  is the False Negative rate, and  $TN$  is the True Negative rate.

Another common metric is the F1-Measure or F1 score which is the harmonic mean of the precision and recall and which could be used when a balance between precision and recall is needed and when the class distribution is uneven (i.e. high  $TN + FP$ ). F1 score is defined as follows:

$$\textit{F1-Measure} = 2 \frac{P * R}{P + R}. \quad (11)$$

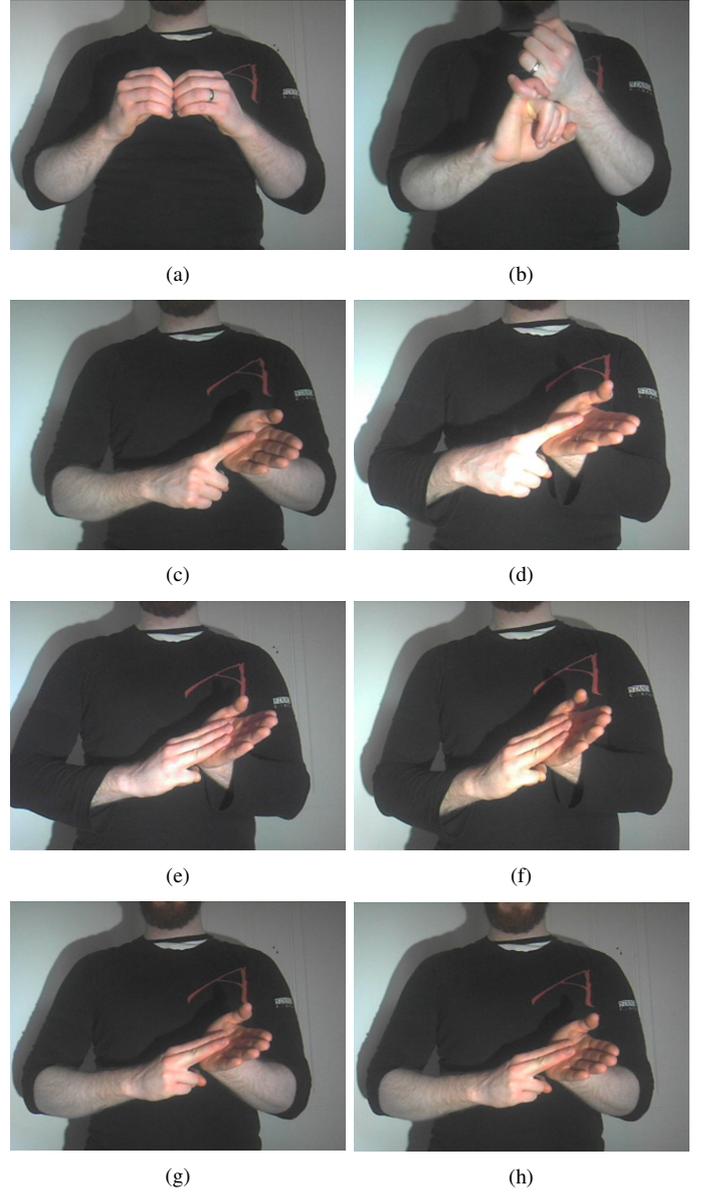


Fig. 7. Samples of our system performing successfully BSL sign recognition of the (a) B letter; (b) S letter; (c)-(d) L letter, in images with different illumination conditions; (e)-(f) M letter in images with different yaw rotations; (g)-(h) N letter in images with different pitch rotations.

As BSL fingerspelling datasets are difficult to find openly, a BSL dataset was created using the dataset creator application developed for this project.

This dataset is unique in the study, as it comprehends BSL phonemes consisting of two-handed gestures. Dataset samples are shown in Fig. 7. This BSL dataset contains 2,600 images of sign classes from  $A$  to  $Z$ . It is worth noting that  $H$  and  $J$  letter images were taken with last endpoint of motion gesture as an approximation. The images are in the Portable Network Graphics (.png) format and have a 640x480 size, a resolution of 72dpi, and a bit depth of 24. Moreover, the lighting is varying, and it also contains an even 50% mixture of signs



Fig. 8. Samples of our system performing hand detection in images with (a) cluttered background; (b) patterned background; (c) high luminance; or (d) high exposure.

with sleeves up and down, in order to introduce variation to the training model.

The dataset has been created in Python language, using the *Dataset Creator* application. The Python language has been chosen due to its lightweight nature and portability, but also because OpenCV wrappers are available for it. Indeed, the OpenCV 'Cv2' library for Python is used to access the webcam data when creating the image dataset and is therefore well suited to the integration with the rest of the project applications.

The prototype is written in Java and uses OpenCV 3.4.3. as well as the Java Native Interface (JNI) platform to enable Java running on the Java Virtual Machine (JVM) to interact with native platform applications written in lower-level languages such as C++. Hence, this JNI interface for the OpenCV dependency provides a critical interface into low-level hardware operations to run OpenCV processes through the C++ library. This interface enables thus the more computationally expensive parts of the application to run more efficiently and thence, provides a better user experience on the low-end phone hardware.

The training of the classifier has been performed using 520 samples (i.e. 26 classes with 20 samples per class) on a computer with features such as AMD FX-8320 3.5GHz (8 cores, 8 threads), 32nm architecture, 8Gb dual-channel DDR3 @ 802MHz, Windows 10 Home, 931Gb Western Digital SATA (7.2k rpm).

For the training function of the Offline Trainer, the datasets used the same imaging kernel as our smartphone application and they were processed image by image using a recursive Image Runner class; this Image Runner taking an input directory and processing each image in the parent folder and any subfolder.

The testing of the prototype has been run on a ZTE Blade V7 smartphone. This is a low-end budget Android phone (sub-£100) for the purposes of encouraging efficient programming and aiding in the overall availability of the end product. This phone model's specifications are as follows: Android 6.0 (Marshmallow) OS, Chipset Mediatek MT6753 (28nm), Octa-core 1.3 GHz Cortex-A53 CPU, Mali-T720MP3 GPU, 16 GB, 2 GB RAM of internal memory, and camera features such as 13 MP, PDAF, Dual-LED dual-tone flash, HDR, panorama, with 1080p @30fps video recorder.

The testing function of the offline trainer is entirely automated. After a training run is complete, the associated test

data is run through the previously trained model. These test images have been subtracted from the initial training dataset in order to not render the testing redundant. Each test image is classified with a predicted letter, and then evaluated against the actual class of the input image. With this data, the outcome of the tests in terms of True Positive ( $TP$ ), True Negative ( $TN$ ), False Positive ( $FP$ ), and False Negative ( $FN$ ) rates are computed. The output of these test operations goes to result log files for analysis and computation of Eqs. 7-11 for each run. Average recognition results can be found in Tables 1-2.

Tests have been carried out in the wild and included 13,066 sample images. BSL uses mainly two hands to represent a letter of the alphabet. However, the letter 'C' is signed using one hand only. Thence, we performed tests for both one and two hands, as illustrated in Fig. 8.

We reported in Tables 1-2 the primary data obtained by processing the approach presented in [21] and our method, respectively, on the BSL large-scale dataset. The approach of [21] involves visual features such as Edge Orientation Histogram features (EOH), whereas our system uses Histogram of Oriented Gradients (HOG) features; the classifier being the Support Vector Machine (SVM), with a linear and a radial basis function (RBF) kernel, respectively.

We can observe that our method combining HOG features and the SVM classifier with a RBF kernel outperforms the state-of-the art approaches in regards of both recognition accuracy and computational efficiency, while performing in the wild.

Furthermore, our method has been compared to available secondary data in the literature. The work of [11] studies Genetic Algorithms (GA) as classifiers for gesture recognition (using only 6 different gesture classes), achieving 98.6% accuracy, but reaching time frameworks in the range of dozens of seconds for the overall image processing. Moreover, [11] requires hundreds of samples for training and has only been tested on 100 samples.

On the other hand, [8] uses a Hidden Markov Model (HMM) and has a BSL recognition accuracy rate per letter of 84.1%, whereas our BSL recognition accuracy rate per letter is 99% and our processing time of 170ms (i.e. less than the 400ms mentioned in the study of [14]) ensures a real-time visual sign recognition pace.

TABLE I  
SIGN RECOGNITION PERFORMANCE USING DIFFERENT METHODS.

Method	Features	SVM Kernel	Precision	Recall	Specificity	Accuracy	F1-Measure
[21]	EOH	Linear	0.853	0.867	0.994	0.988	0.852
[21]	EOH	RBF	0.855	0.868	0.994	0.988	0.854
Our	HOG	Linear	0.863	0.874	0.994	0.989	0.861
<b>Our</b>	<b>HOG</b>	<b>RBF</b>	<b>0.869</b>	<b>0.880</b>	<b>0.995</b>	<b>0.990</b>	<b>0.867</b>

TABLE II  
AVERAGE PROCESSING TIME OF THE DIFFERENT METHODS PERFORMING SIGN RECOGNITION.

Method	Features	SVM Kernel	Average Processing Time (s)
[21]	EOH	Linear	0.178
[21]	EOH	RBF	0.178
Our	HOG	Linear	0.173
<b>Our</b>	<b>HOG</b>	<b>RBF</b>	<b>0.170</b>

#### IV. CONCLUSIONS

The paper proposes an assistive technology performing British Sign Language (BSL) alphabet translation in real-time and in real-world conditions, with an accuracy of over 99%. The design aims to provide an inclusive and accessible solution consisting in an intelligent vision system for automated BSL fingerspelling recognition, without being invasive or financially expensive. The algorithms developed within this system include Histogram of Oriented Gradients (HOG) method and the Support Vector Machine (SVM) technique. The resulting smartphone application has been successfully tested on a large-scale dataset in the wild. Its excellent performance leads, on one hand, to an accessible, assistive HCI product for non-deaf people wishing to learn BSL and/or to communicate using BSL with deaf persons, and on the other hand, to a potential HRI product for companion robots having the task to assist hearing and/or speech impaired people.

#### REFERENCES

- [1] Actiononhearingloss.org.uk, "Facts and Figures,," 2019, Available online at: <https://www.actiononhearingloss.org.uk>.
- [2] BDA, "British Deaf Association,," 2019, Available online at: <https://bda.org.uk>.
- [3] NRCPD, "The National Registers of Communication Professionals working with Deaf and Deafblind People,," 2019, Available online at: <https://www.nrcpd.org.uk>.
- [4] D. Waters, R. Campbell, C.M. Capek, B. Woll, A.S. David, P.K. McGuire, M.J. Brammer, and M. MacSweeney, "Fingerspelling, signed language, text and picture processing in deaf-native signers: The role of the mid-fusiform gyrus," *NeuroImage*, vol. 35, no. 3, pp. 1287–1302, 2007.
- [5] J.I. Olszewska, "Designing transparent and autonomous intelligent vision systems," in *Proceedings of the International Conference on Agents and Artificial Intelligence (ICAART)*, 2019, pp. 850–856.
- [6] G. Plouffe and A.-M. Cretu, "Static and dynamic hand gesture recognition in depth data using dynamic time warping," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 2, pp. 305–316, 2016.
- [7] M. Goyal, B. Shahi, K.V. Prema, and N.V.S.S. Reddy, "Performance analysis of human gesture recognition techniques," in *Proceedings of the IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology*, 2017, pp. 111–115.
- [8] S. Liwicki and M. Everingham, "Automatic recognition of fingerspelled words in British sign language," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 50–57.
- [9] J.L. Raheja, A. Mishra, and A. Chaudhary, "Indian sign language recognition using SVM," *Pattern Recognition and Image Analysis*, vol. 26, no. 2, pp. 434–441, 2016.
- [10] A.B. Jani, N.A. Kotak, and A.K. Roy, "Sensor based hand gesture recognition system for English alphabets used in sign language of deaf-mute people," in *IEEE SENSORS Proceedings*, 2018, pp. 1–4.
- [11] D.-J. Li an Y.-Y. Li, J.-X. Li, and Y. Fu, "Gesture recognition based on BP neural network improved by chaotic genetic algorithm," *International Journal of Automation and Computing*, vol. 15, no. 3, pp. 267–276, 2018.
- [12] S. Salian, I. Dokare, D. Serai, A. Suresh, and P. Ganorkar, "Proposed system for sign language recognition," in *Proceedings of the IEEE Conference on Computation of Power, Energy Information and Communication*, 2017, pp. 58–62.
- [13] B.L. Loeding, S. Sarkar, A. Parashar, and A.I. Karshmer, "Progress in automated computer recognition of sign language," in *Proceedings of the International Conference on Computers for Handicapped Persons*, 2004, pp. 1079–1087, LNCS, Springer.
- [14] E. Klima and U. Bellugi, *The Signs of Language*, Harvard University Press, 1979.
- [15] M. Loesdau, S. Chabrier, and A. Gabillon, "Hue and saturation in the RGB color space," in *Proceedings of the International Conference on Image and Signal Processing*, 2014, vol. 8509, pp. 203–212, LNCS, Springer.
- [16] C. Rouge, S. Shaikh, and J.I. Olszewska, "HD: Efficient Hand Detection and Tracking," in *Proceedings of the IEEE Federated Conference on Computer Science and Information Systems (FedCSIS)*, 2016, pp. 291–297.
- [17] J.I. Olszewska, C. De Vleeschouwer, and B. Macq, "Multi-feature vector flow for active contour tracking," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2008, pp. 721–724.
- [18] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005, pp. 886–893.
- [19] C. Cortes and V.N. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [20] P. Matykiewicz and J. Pestian, "Effect of small sample size on text categorization with support vector machines," in *Proceedings of the ACL Workshop on Biomedical Natural Language Processing (BioNLP)*, 2012, pp. 193–201.
- [21] S. Nagarajan and T.S. Subashini, "Static hand gesture recognition for sign language alphabets using edge oriented histogram and multi-class SVM," *International Journal of Computer Applications*, vol. 82, no. 4, pp. 28–35, 2013.

# Counting Instances of Objects in Color Images Using U-Net Network on Example of Honey Bees

Weronika W. Westwańska

Zespół Szkół Stowarzyszenia Rodzin  
Katolickich Archidiecezji Katowickiej  
im. Kardynała Prymasa Augusta Hlonda  
ul. Kościuszki 11, 41-500 Chorzów, Poland  
Email:  
mkw.weronika.westwanska@gmail.com

Jerzy S. Respondek

Silesian University of Technology,  
Institute of Informatics, AEI Faculty,  
ul. Akademicka 16, 44-100 Gliwice, Poland  
Email: jerzy.respondek@polsl.pl

**Abstract**—This article presents a novel approach to segmentation and counting of objects in color digital images. The objects belong to a certain class, which in this case are honey bees. The authors briefly present existing approaches which use Convolutional Neural Networks to solve the problem of image segmentation and object recognition. The focus however is on application of U-Net convolutional neural network in an environment where knowledge about the object of interest is only limited to its rough, single pixel location. The authors provide full access to the details of the code used to implement the algorithms, as well as the data sets used and results obtained. The results show an encouraging low level of counting error at 14.27% for the best experiment.

## I. INTRODUCTION

CONVOLUTIONAL Neural Networks (CNNs) are considered state of the art architectures for object detection and segmentation in color images [1]-[4]. In this article we apply a specific type of CNN, called U-Net CNN (UNETCNN) [5], [6], to count instances of honey bees in color images captured by digital video camera. The dataset was sourced from [7] and is freely available to anyone who wishes to test their own bee counting routines. We are proposing a novel way of preparing data modeling for a UNETCNN, where the only information available, about the object of interest (OOI), is its approximate location, defined as a single point in two dimensional planes. Majority of research in the area of CNNs assumes that an OOI location is provided by a rectangle tightly encompassing its border [4]. We propose to use a circle with its center placed on the OOI. The assumption is: we are not focusing on finding exact boundaries of OOI, but rather on counting instances of the OOI, versus manually provided data in a segmentation set. As we are going to show in this paper, it is not necessary to cover an OOI with a bounding box, to achieve high classification accuracy. Instead we are only considering a location of a pixel lying on the surface of an

OOI. The location of the pixel constitutes a center of a small circle, which contains pixels belonging mostly to OOI and partially to the background. This step is called UNETCNN data generation. Following that, we employ further steps: training of UNETCNN, using the trained model for automatic segmentation of OOI, automatic counting of OOI instances. The final step allows computation of the relative error by comparing the number of OOI instances that were detected automatically versus how many of them were manually labelled by a human.

The article is organized as follows. In section II we discuss recent works in the area of adopting neural network to solve the problem of speed and accuracy in image recognition process including bee detection. We also present UNETCNN architecture and give reasons for adapting it for our own solution. Section III describes in details each stage of our experiment and presents the results. In section IV we summarize our work and discuss future directions which could lead to interesting findings.

## II. EXISTING APPROACHES TO BEE RECOGNITION PROBLEM AND UNET CHARACTERISTIC

### A. Different approaches to OOI detection

The bee detection problem was analyzed in [8] with various scenarios according to diversity of background characteristic, light intensity, bee size, image segmentation and labelling efficiency etc. The experiment evaluation helped us to decide which aspects of image recognition are the most important in our experiment and suggested the way of training set preparation. It also convinced us that Adam optimizer [9] is a good choice for training our neural network.

In [10] bees' recognition problem is discussed in the field of different methods of object recognition. CNNs are compared with Multi-Layer Perceptron (MLP) models. According to the experiments and results presented in that

This work was supported by Statutory Research funds of Institute of Informatics, Silesian University of Technology, Gliwice, Poland (BK/204/RAU2/2019).

paper, it was found that MLP performance is much worse than CNN (taking into account the same dataset). The author stated that ADAM optimizer [9] gave reliable results in comparison to others. In terms of kernel size it was suggested that choosing it as 5x5 pixels, provides better performance of the model. In [10] it was noticed, that one of the problems in proper classifying the bees is a possible presence of bee shadow, which has got the same shape as the bee.

### B. UNet

UNETCNN is a NN architecture designed for image segmentation, characterized by low demand on number of annotated training images, and fast data processing. In [5] the authors use medical images as source of training data and demonstrate state of the art results, compared to manual labelling, making this architecture a de facto standard in medical images segmentation.

A typical UNETCNN consists of two paths. One (the contracting path) is represented by a typical CNN with two operations of convolution and max pooling following one after another. This path reduces spatial information, but provides better information where OOI might be present. The second path (expansive) matches the features extracted in contracting path using a sequence of up-scaling transformations.

A game changing innovation in [5] was the fact that a small set of training images can yield more precise segmentations than larger training sets in other algorithms. For a UNETCNN the training data is sourced from the images by dividing them into smaller windows, and then randomly chosen to be included in a training set.

Another feature of UNETCNN is that the algorithm enables finding the solution not only for diverse data set, but also for relatively similar data. Normally lack of diversity would lead to difficulties in recognizing objects in images which do not present features close to the ones the network was trained on. In order to alleviate this phenomenon and make the results independent of the input data, an excessive data augmentation is used. The network gains the data not only straight from the input data, but also from elastic deformations of the training images. This way the network training process can be invariant to deformations even if the images used in the process do not contain enough OOI. Recent works on UNETCNN [6], [11]–[14] prove that this architecture yields very good results and currently might be the best for solving objects recognition problem not only for 2D but also for 3D data.

In our approach we decided not to segment the training data manually, because of the amount of time it would take. We decided to verify if it would be practical to find a way of solving the problem of manual segmentation, by not providing bounding box or a mask for each object

in the image. We decided to segment images based only on single points for each OOI located in any of the training images.

## III. THE PROPOSED SOLUTION

As we mentioned above, there are 4 stages applied in OOI counting, each described in detail below. For the purpose of our experiments, we decided to use data set downloaded from [7]. At the time of performing experiments, the dataset contained only 550 manually labelled images. We enhanced this data set with a further 1086 images, manually labelled by us, using custom written software. The algorithms examined in this work are publicly available via [15].

### A. Stage 1 – UNETCNN input data generation

As it was commented earlier we decided to adopt a shape of a circle to describe part of an area belonging to each OOI. As the input images are of size 640 x 480 pixels we empirically set the size of the circle to be 16 pixels in radius (which is fully configurable). The idea is that when an OOI is labelled, a point is placed on its surface. We assumed that the pixels lying within the nearest neighborhood of the labelled location belong with a high probability to the OOI itself. We decided to simulate such a neighborhood with a circle of a chosen radius, where the pixels lying more towards the circle's edges are less likely to be part of the OOI itself. A linear probability decrease function is implemented with a minimum probability  $P_{min}$  declared for the edge of the circle and maximum probability  $P_{max}$  for the center of the circle. We set these values to 0.99 and 1.0 respectively for circles with radius of 16 pixels, and to 0.80 and 1.0 for circles with radius of 20 pixels (within different sets of experiments – see Table I). Any pixels lying outside of the circle are considered to belong fully to the background. This approach means that the problem becomes a binary classification one, where we set class 0 as background, and class 1 as an OOI. Any pixel in the generated modelling data has a probability associated with it: background and foreground. Both probabilities sum up to 1.0.

After all the images from the dataset were labelled, and the parameters of the data generation decided (such as circle radius,  $P_{min}$  and  $P_{max}$ ), we could finally create a numeric representation of the modelling data, stored as two separate Python numpy files. The first file described the color RGB channels of the pixels from the labelled images, normalized to values between 0.0 and 1.0. The second file described probability values for non OOI (NOOI) and OOI classes for every pixel, meaning that each pixel has a 2 dimensional feature vector associated with it.

The idea behind such generalization for the shape of an OOI as a circle, was that we were dealing with bees, which are relatively similar in size. We also were using the

fact that NN is working with fuzzy data, where we can assume that a pixel can partially belong to the NOOI and partially to the OOI. Based on that assumption and the fact that training algorithms for NN find local optimal solution we decided to test if this approach worked.

### B. Stage 2 – Training UNETCNN model

In previous stage 1 we created 2 sets: one representing samples X, and the second set Y representing corresponding values mapped to probabilities of all the pixels from modelling set. These sets stored on a hard drive are quite large in size, depending on amount of images in the modeling set. The X and Y sets, stored as Python numpy files, were used as a source for training of the UNETCNN model. The model we used was a slight modification from the original, with last layers changed to use 2 instances of RELU layer, followed by a Dropout layer and SoftMax used for classification. This solution was introduced to minimize the risk of overfitting the network and gave significant improvement according to results of computations. SoftMax layer enabled classification of background and foreground classes and would allow us to use it for classification of more than only two classes in the future.

In order to provide the training data for the UNETCNN, the generated data had to be randomly accessed to retrieve rectangular windows of pixels which contained modelling samples for the OOI and the NOOI classes. We adopted an algorithm, where each modeling image has a total of 80 (configurable value) windows randomly retrieved. Among the 80 windows, a specific amount of windows are considered as OOI windows and the rest are considered as NOOI windows. In order for a window to be OOI based, it has to pass a minimum percentage threshold for OOI pixels count. The amount of OOI windows would vary per image, depending on how many manually labelled OOI samples were present, versus total OOI samples in the modelling set. For example if image IMG1 had 2 OOIs labelled, and image IMG2 had 4 OOIs labelled, then there would have to be 2 times less OOI based windows randomly selected from IMG1 than compared with a number of randomly selected OOI based windows from IMG2. This individual approach was dictated to preserve a balanced number of OOI and NOOI classes in training data. The details of how the modelling set is created from the data generated in Stage 1 are available in [15].

The modelling data collected so far was then further split into training and validation sets (80/20 ratio) to be used in UNETCNN. On average we achieved a decent 96% validation set accuracy.

### C. Stage 3 – OOI Segmentation

After the UNETCNN model was trained we could use it to perform OOI detection on images from segmentation set. As it was mentioned in Stage 2, the model operates on assumption that the input tensor used in classification is of certain dimensions. In our experiments the dimensions were  $N \times 32 \times 32 \times 3$ , where N is the number of windows collected from the segmented image,  $32 \times 32$  are width and height of the window, and 3 stands for RGB channels normalized to [0,1] interval. The value of 32 is also configurable within the source code, and corresponds to diameter of the OOI modelling circle from Stage 1 [15].

We propose a custom approach for segmenting an input image, where a set of windows meshes is created. A single mesh is started at a specific (x, y) offset from the top left corner of a segmented image. The idea is to cover as much of the image as possible with windows tightly attached to each other, where every window needs to be wholly fitting into the image. The windows and their pixel RGB values would create a set which is then classified by the trained UNETCNN model from stage 2. The results of each mesh's classification were then added into a special matrix with values taken for OOI class at corresponding locations to the mesh pixels. Another matrix is also kept to count how many classifications were computed for each pixel from the segmented image. In the end, when all meshes are classified, the accrued classification results are scaled (using the accumulated classification counts), so that a heat map is created. The meshes created for the segmentation process are generated at a specified step of 2 pixels (configurable value) from coordinates of (0, 0) to (32, 32), where 32 is a size of classification window for UNETCNN.

The fore mentioned heat map is later converted to a binary image, where each pixel is decided to be as a part of the OOI or part of the background, based on the scaled voting produced by meshes.

### D. Experiments - Counting OOI

In our experiments we decided to examine how different parameters affect counting error which is expressed as a difference between 100% and a percentage of automatically detected instances of OOI (bees) versus total amount of manually labelled instance of OOI. In Table I we present OOI relative counting error, dependent on modelling size set and window size used for training of U-Net and further segmentation. The error progression is visualized in Fig. 1. Minimum percentage of pixels per window, so that can be considered as an OOI based window, was set as 45 and 50 for Experiment 1 and Experiment 2 respectively. More details can be found in logs provided in [15]. It turns out that the more images are available, the better results. Interestingly, error started dropping dramatically at about 500 images, reaching its

minimum value for full modelling set size of 1096 images with the OOI window size set at 40 pixels (Experiment 1).

TABLE I.

BEES COUNTING ERROR DEPENDING ON AMOUNT OF IMAGES USED IN MODELING SET ALONG WITH WINDOW SIZE

Attempt number	[Modeling images count, window size]	OOI Relative Counting Error
1	[100,40]	75.08%
2	[200, 40]	73.00%
3	[500, 40]	24.23%
4	[1000, 40]	16.83%
5	[1096, 40]	14.27%
6	[100, 32]	75.24%
7	[200, 32]	74.55%
8	[500, 32]	26.35%
9	[1000, 32]	16.87%
10	[1096, 32]	18.17%

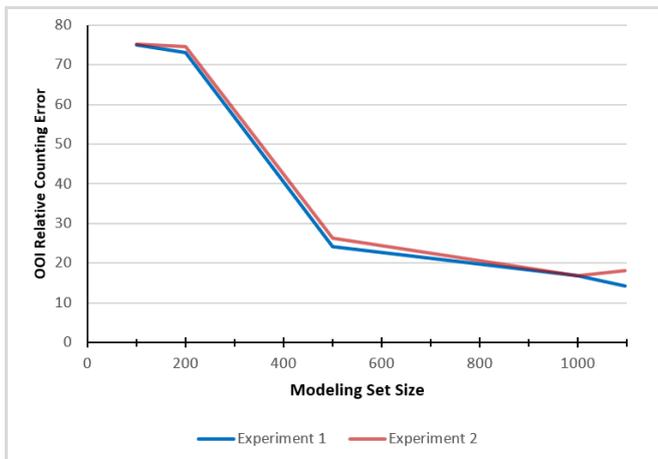


Fig. 1 OOI Counting Relative Error for experiments on parameters set 1 and set 2.

#### IV. CONCLUSIONS

In this article we presented a review of the most recent methods of OOI segmentation and detection. Based on these we chose a UNETCNN architecture which we adapted to a fairly new topic of counting of OOI, based only on their singular locations in the training set. We developed a new approach for generating modeling data, using the trained UNETCNN for segmentation, and further for counting occurrences of OOI in the validation set. We reached a satisfactory level of error at 14.27%, which encourages us to pursue this topic further. The experiments performed show that the training data preparation does not have to be mundane and time consuming and just a few hours spent on the process of labeling images can yield good results, not only in terms of counting error reduction

but also decent outcome in OOI segmentation. The Python code which was developed for the purpose of this research is available freely to anyone from [15]. We would like to thank Mr. Jonathan Byrne for making his data set available in [7].

#### REFERENCES

- [1] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, Cambridge CA: Massachusetts, pp. 321–359, 2016. <https://www.deeplearningbook.org/>
- [2] E.R. Davies, *Computer Vision. Principles, Algorithms, Applications, Learning*, 5th ed., London, pp. 456–462, 2018.
- [3] R. Yamashita, M. Nishio, R. Kinoshita, K. Togashi, "Convolutional neural networks: an overview and application in radiology", *Insights into Imaging*, vol. 9, pp. 611–629, 2018. <https://doi.org/10.1007/s13244-018-0639-9>
- [4] Z. Zhao, P. Zheng, S. Xu, X. Wu, "Object detection with deep learning: A Review", *Journal of LaTeX Class Files*, vol. 14, no. 8, 2017. <https://doi.org/10.1109/TNNLS.2018.2876865>
- [5] O. Ronneberger, P. Fischer, T. Brox, "U-Net: convolutional networks for biomedical image segmentation", *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, 2015. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
- [6] K. H. Jin, M. T. McCann, E. Froustey and M. Unser, "Deep convolutional neural network for inverse problems in imaging", *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4509–4522, 2017. <https://doi.org/10.1109/TIP.2017.2713099>
- [7] <https://www.kaggle.com/jonathanbyrne/to-bee-or-not-to-bee>, accessed on the 1st of February 2019.
- [8] M. Kelcey, "Counting bees on a rasp pi with a conv net", 2018. [http://matpalm.com/blog/counting\\_bees](http://matpalm.com/blog/counting_bees)
- [9] P. Kingma, J. Lei Ba, "ADAM: a method for stochastic optimization", *arXiv preprint arXiv:1412.6980*, 2014. <https://arxiv.org/abs/1412.6980>
- [10] A. Tiwari, "A deep learning approach to recognizing bees in video analysis of bee traffic", *Utah State University All Graduate Theses and Dissertations*, 7076, 2018. <https://digitalcommons.usu.edu/etd/7076/>
- [11] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation", *Medical Image Computing and Computer-Assisted Intervention*, vol. 9901, pp. 424–432, 2016. [https://doi.org/10.1007/978-3-319-46723-8\\_49](https://doi.org/10.1007/978-3-319-46723-8_49)
- [12] J. Chen, L. Yang, Y. Zhang, M. Alber, D.Z. Chen, "Combining Fully Convolutional and Recurrent Neural Networks for 3D Biomedical Image Segmentation", *NIPS'16 Proceedings of the 30th International Conference on Neural Information Processing Systems*, pp. 3044–3052, 2016. <https://arxiv.org/abs/1609.01006>
- [13] J.P. Viguera-Guillén, B. Sari, S.F. Goes, H.G. Lemij, J. van Rooij, K.A. Vermeer, L.J. van Vliet, "Fully convolutional architecture vs sliding-window CNN for corneal endothelium cell segmentation", *BMC Biomedical Engineering*, vol. 1, 2019. <https://doi.org/10.1186/s42490-019-0003-2>
- [14] S. Baek, Y. He, B.G. Allen, J.M. Buatti, B.J. Smith, K. A. Plichta, et al. "What does AI see? Deep segmentation networks discover biomarkers for lung cancer survival", 2019. <https://arxiv.org/abs/1903.11593>
- [15] <https://github.com/WeronikaWestwanska/ToBeOrNotToBee>, accessed on the 8th of May 2019.

# Generating Human Mobility Route Based on Generative Adversarial Network

Ha Yoon Song  
Department of Computer Engineering  
Hongik University  
Seoul, Republic of Korea  
Email: hayoon@hongik.ac.kr

Moo Sang Baek  
Research Institute of  
Science and Technology  
Hongik University  
Seoul, Republic of Korea  
Email: moosangbaek@gmail.com

Minsuk Sung  
Department of Computer Engineering  
Hongik University  
Seoul, Republic of Korea  
Email: mssung94@mail.hongik.ac.kr

**Abstract**—Recently, many researches on human mobility are aiming to suggest the personal customized solution in the diverse field, usually by academia and industry. Combined with deep learning methods, it is able to predict and generate novel routes of objects from the mobility data including the given past trends. In this work, Generative Adversarial Network (GAN) model is introduced for creating individual mobility routes based on sets of accumulated personal mobility data. The mobility data had been collected by use of geopositioning system and personal mobile devices. GAN has Discriminator and Generator which are composed of neural networks, and can train and extract geopositioning information. A sequence of longitude and latitude can be geographically mapped, and matrices including all these information can be handled by GAN. The GAN-based model successfully handled individual mobility routes in this way. Consequently, our model can generate and suggest unexplored routes from the existing sets of personal geolocation data.

## I. INTRODUCTION

INDIVIDUAL mobility data has a huge capacity and can provide significant information and knowledge to modern industries. By processing the result of analyzed mobile dataset, enterprises can get human personalities so that they can interact with their customers effectively and set directions in the field of marketing. For instance, the analysis of the consumers' data can help to determine the location of new commercial shop branches and to find the intersected location where their customers visit in common. Therefore, the companies can provide service improving customer satisfaction through the individual preference which is extracted from correlation among distinguished locations analyzed from mobility data.

Normally, personal location data can be divided into location clusters according to the data distribution, and movement patterns in each cluster vary depending on the individual's purpose and desire. It means that the location clusters imply individual lifestyles. By analyzing the correlation between the major location clusters, we could deduce the past visited points and predict the future visitations. This correlation is an important factor which is necessary for creating a new mobility route. Indeed, an appropriate data pre-processing is

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (NRF-2019R1F1A1056123).

required for training GAN with human location data. In data pre-processing work, we have partitioned personal mobility data, and utilized 4 layers of Convolutional Neural Networks (CNNs) to deliver features of the partitioned data. In data post-processing work, we adopted CNN to preserve essential location feature in the form of latitude and longitude by improving GAN performance. We introduce GAN for generating mobility routes. In this GAN, there are two representative Discriminator and Generator networks. These networks compare geopositioning features extracted from each mobility data with the generated probability distributions, while backpropagating the differences in each network. Discriminator learns the features extracted from the whole individual mobility route with extra mobility route, and improves ability to discriminate the fake route obtained from Generator and the real route from input dataset. Consequently, the self-creating GAN network, which ensures to enhance the discriminative features, can generate unexplored routes for area size of 3 square kilometer.

Section II will discuss the related researches on human mobility data with diverse approaches. In section III, we analysis our raw data, and section IV explains method we use on the data pre-processing and post-processing. Section V describes detailed methodology with proposed GAN. In section VI presents the result of generation route with our model. Section VII concludes our research and discusses about our future research.

## II. RELATED WORKS

In this section, we review previous investigations on human mobility patterns, which were focused on creating novel mobility route or next location by data mining. It is a typical approach to predict the next location of objects through mobility sequence tree generation by pattern mining the mobility of the object (Pfoser et al. 2000 ; Ying et al. 2011) [1] [2]. A further approach suggests the mobility tree which is expanded gradually by pattern mining (Gorawski and Jureczek. 2010) [3], and there is utilization with Location Based Service(LBS) to generate mobility patterns (Lee et al. 2004) [4]. Through such various methods, including mobility pattern tree generation and next location prediction, the combination of the probabilistic approach and data mining techniques has recently been

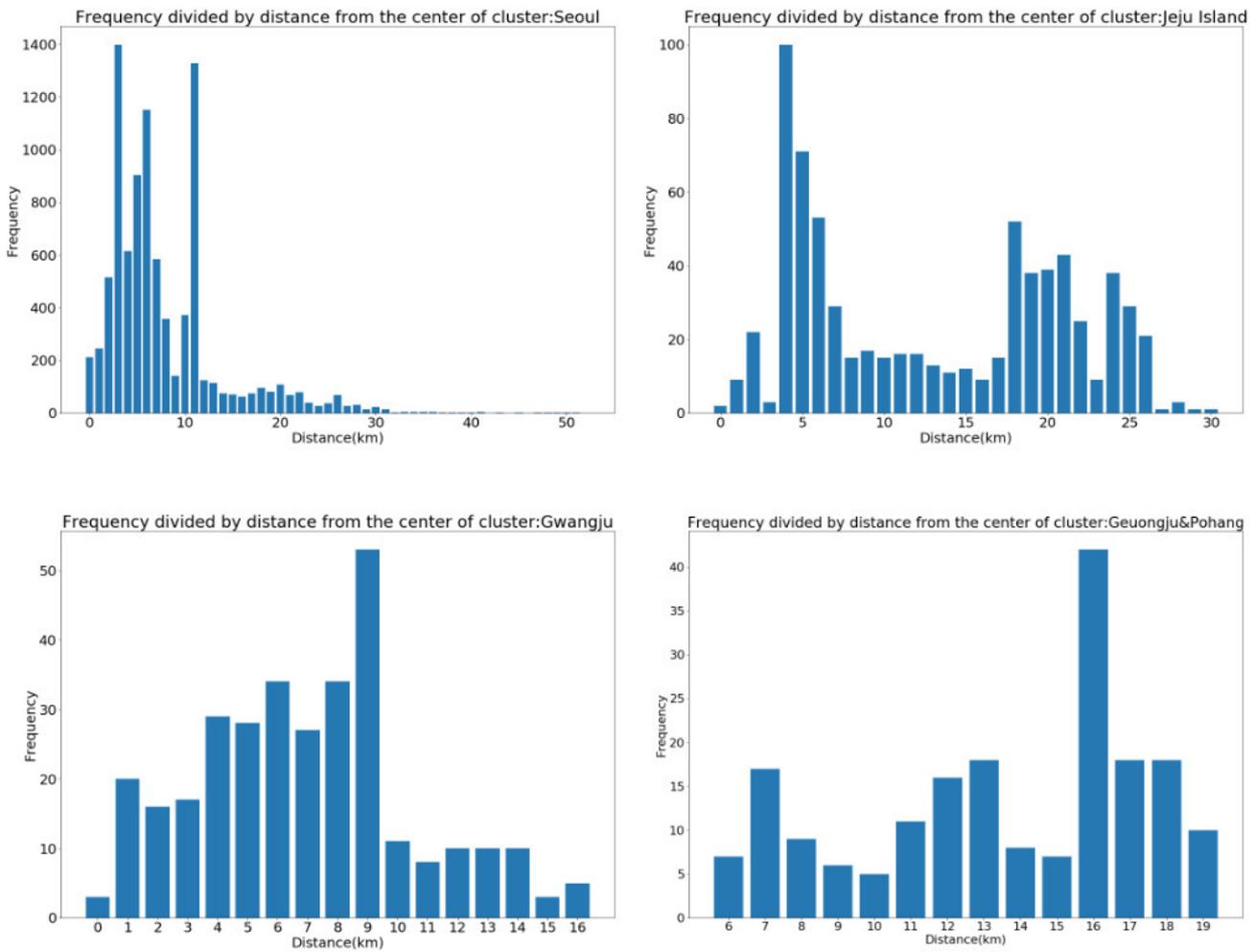


Fig. 1. Location Distribution inside Clusters

applied (Monreale et al. 2009 ; Jeung et al. 2008) [5] [6]. The graphs generated from the data mining results of accumulated trajectory patterns predict the next location (Morzy. 2006 ; Morzy. 2007) [7] [8]. Trajectory pattern can be utilized to mining past trajectories without trees, graphs or probability models (Giannotti et al. 2007) [9]. The success of combination of frequent trajectory and mobility rule for location prediction (Nhan and Ryu. 2006, Song and Choi. 2015) [10] [11] assumes diverse situations, such as disaster from the perspective of location prediction(Sudo et al. 2016) [12]. Markov chain-based approach has been applied on human mobility pattern and achieved remarkable performance on the next location prediction based on human mobility pattern(Baratchi et al. 2014) [13]. In the previous period, researchers have applied the data mining, trajectory pattern tree and Markov chain tools mainly.

From the view of GAN, GAN has achieved impressive outcomes in image generation (Alec Radford, Luke Metz, and Soumith Chintala. 2016) [14], image translation (Yunjey Choi

et al. 2018) [15], and there have been several recent researches to analysis synthetic data generation. Research regarding the activity patterns of neurons using GAN to generate the data can be found in (Molano-Mazon et al. 2018) [16]. Furthermore, movement trajectories based on socially acceptable behavior had been investigated as shown in (Gupta et al. 2018) [17]. Behavior includes passing or meeting of people during walking with parameters such as speed and direction. Current researchers for deep data generation on human mobility are (Alzantot, Chakraborty, and Srivastava. 2017) [18], using an architecture similar with GAN model, which is mixture of density networks to generate the acceleration time series data. However, this research did not use complete GAN architecture, and the input dataset of this research was not geopositioning data of trajectories but solely acceleration data. The purpose of our research is to generate novel mobility routes based on a plenty of daily mobility patterns.

TABLE I  
BASIC PROPERTIES OF LOCATION CLUSTERS

Cluster No.	Location	Center of Cluster(Latitude / Longitude)	Total Number of Data in Cluster
Cluster 0	Seoul	33.48852999 126.47904523	9668
Cluster 1	Jeju Island	33.44230848 126.52364544	728
Cluster 2	Gwangju	35.16296341 126.88780208	317
Cluster 3	KyeongJu, Pohang	35.93293456 129.31300085	191

### III. GEOPOSITIONING DATA ANALYSIS

The mobility data collected by positioning devices, such as smart phones, have the latitude, longitude and time information. The raw data we use are containing the location information of a specific object for more than three years.

#### A. Macroscopic View

In analyzing data from a macro perspective, the K-means clustering method is applied to extract largely four clusters based on location and density from the collected data distribution [19]. The basic properties of each location cluster are shown in table I. Fig. 1 shows histograms of the location points of each cluster. The frequency values are shown according to the distance between a center and positioning data in each location cluster [20]. Cluster 0 contains a relatively large number of location data, and this cluster reflects the location data in lifestyle. On the other hand, a low number of data is collected at cluster 1, 2, 3, and its result shows the irregular mobility pattern of object, such as a trip. This implies that the micro-mobility of an object indicates remarkable differences in each cluster depending on the purpose to visit.

#### B. Microscopic View

Understanding correlation between visited points is a starting point for expanding correlation between day-to-day movement trajectories. In cluster 0, there are mobility patterns on the movement of lifestyle, and the mobility patterns consist of visited points in object's daily life. To analyze correlation among distinguished points in cluster 0, it is necessary to extract past visited points. Distinguished points can be found applying K-means clustering algorithm to each day of the three-year full mobility data. Redundancies of 185 geographical points were eliminated, and the entire mobility pattern in cluster 0, made up of each major point, consists of about 2,700 distinguished points.

Learning connection between distinguished points and connection between each movement pattern, where the points are clustered, is identifying the mobility tendency of an object. Our GAN-based model will put daily movement patterns in Discriminator as input dataset, recognizing the existing patterns as real data, and will train dozens of times for Generator to generate mobility route similar to input dataset. This begins with the notion that a mobility route will be created based on object's mobility tendency.

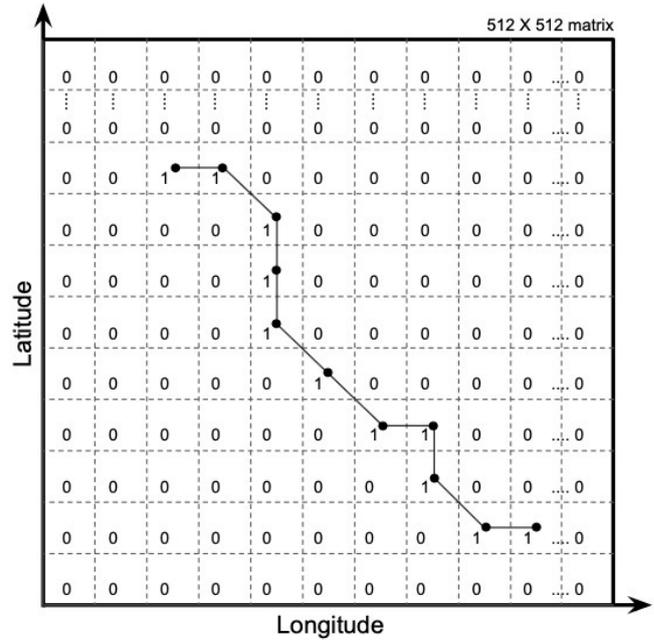


Fig. 2. Area Partitioning and mapping

### IV. GEOPOSITIONING DATA PROCESS

#### A. Data Pre-processing

As the wide range of location dataset is somewhat excessive for GAN model to train, proper data preprocessing is required. The unstructured raw data is initially informal structure, and could not apply to GAN model directly. The area is limited to the size of 3 kilometer in vertical length and 3 kilometer in horizontal length. The input dataset was structured using area partitioning method. The mobility patterns were partitioned as shown in Fig. 2. As a practical approach, area partitioning method can represent daily movement patterns as matrix. In detail, each visited area is mapped from intersection between the trajectory mobility pattern and unit areas. Each unit area size is about 0.06 square kilometer. By the characteristic of the ge positioning data, positioning error, such as value of latitude and longitude, is effectively found from the fifth decimal point of values. That is, eliminating the digits from the 6th to 11th decimal places do not distort the location information. We applied the round function through the area partitioning method. To preserve all positioning information in matrices, the CNN is utilized with 4 layers with each filter size of

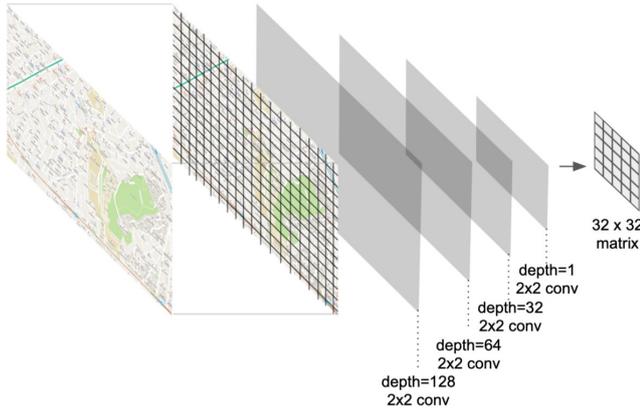


Fig. 3. Process of Convolution

2x2x128, 2x2x64, 2x2x32, and 2x2x1. In data convolution process, input data for GAN model are transformed to 32x32 sized matrices. This problem is solved by making the mobility data to pass multiple CNN layers with Leaky ReLU as an activation function which is convolution stage. Fig. 3 shows the process of convolution. The volume of input dataset was not enough to train the mobility patterns by our model, in detail, the individual accumulated daily mobility pattern. To improve this problem, synthetic inputs, which are extracted by randomly shuffling of 40% to 90% of input data, are duplicated so that GAN model can train various sets of input data. Through this augmentation method, insufficient input dataset amplified 100 times.

### B. Data Post-processing

The output of our GAN model is 32x32 sized matrices. In order to visualize output data on the map, deconvolution as post-processing is required. The CNN as deconvolution layers is used with 4 layers. In this deconvolution process, Nearest Neighbor function is utilized to resize process which transforms the output to 64x64, 128x128, 256x256 and 512x512 size step by step. Fig. 4 shows the steps of deconvolution toward visualization on geographical maps.

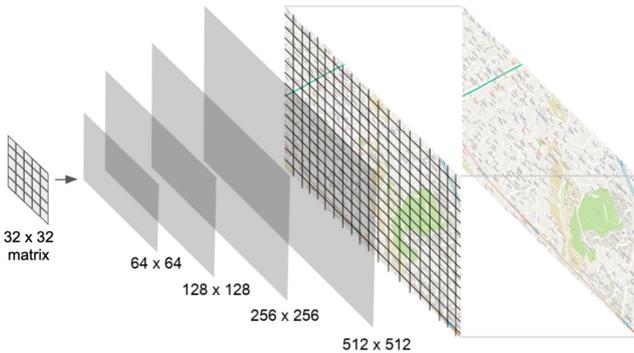


Fig. 4. Process of Deconvolution

## V. METHODOLOGY WITH GAN

GAN was firstly introduced from (Goodfellow et al. 2014) [21], which consists of a Generator and a Discriminator, structured by neural networks. The basic theory of GAN is started below the Non-Saturating game. In detail, the Discriminator minimizes the value of equation 1 and the Generator maximizes the value of equation 2. Generally, GAN model has the saturating issue, which occurs with initialization. Equation 3 will be used as an objective function to solve this issue (Fedus et al. 2017) [22].

$$\max_D \min_G V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(x)} [\log(1 - D(G(z)))] \quad (1)$$

$$\max_D \min_G V(D, G) = E_{z \sim p_z(x)} [\log(1 - D(G(z)))] \quad (2)$$

$$\min_G E_{z \sim p_z(x)} [\log(1 - D(G(z)))] = \max_G E_{z \sim p_z(x)} [\log D(G(z))] \quad (3)$$

We proposed GAN to generate diverse novel mobility routes. In this GAN, the two neural networks compete each other to improve ability to generate mobility route with latitude and longitude features. Discriminator tries to discriminate the real mobility route, training from the accumulated movement trajectory of object and exporting the probability value which is close to “1”, when the input data is distinguished as real data. On the other hand, the Generator has the random latent vector as input, and generates a matrix. Discriminator tries to compare each matrix generated from Generator with input dataset. When it is indistinguishable if the matrix is from Generator or input dataset, the Generator can suggest a novel route which put together a wide range of daily route. Fig. 6 and Fig. 7 shows our GAN architecture for experiment.

## VI. EXPERIMENT

### A. Test of Convolution and Deconvolution

Despite the data processing, we could find that the geopositioning data are preserved when the data are passed through convolution and deconvolution layer. Fig. 8 shows the original route, convolutionized route and deconvolutionized route as diagram. In test of convolution and deconvolution, fig. 9-(a) shows mapped route from raw data, and fig. 9-(b) shows route after deconvolution process without error. From this result, it is appropriate to apply convolution and deconvolution layer with GAN model.

### B. Experiment for Activation Function

In order to figure out the better result, generating route, we experimented with varying activation functions for Generator and Discriminator. Seven activation functions are applied in this work, such as ReLU, ReLU6, Softplus, Tanh, Softsign and eLU (Fig. 5) with 20,000 and 50,000 epoch. Among

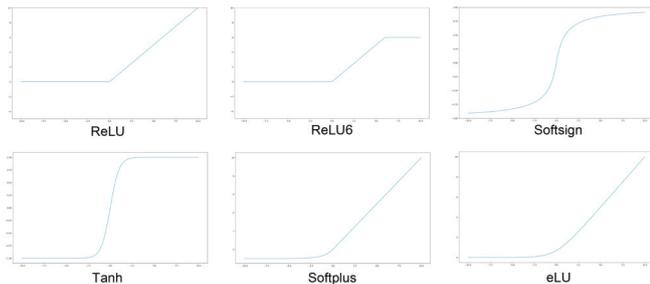


Fig. 5. Activation Functions

the result with the 7 activation functions, ReLU activation function made the worst output, and Tanh activation function made the best output. During training with ReLU, it has possibility to partially miss some positioning data. On the other hand, during training with Tanh, even if Tanh reduce the positioning data, the mobility route can be produced without missing positioning data. This is because the network of the model tends to partially lose its path data as it passes through the ReLU activation function while learning with the input paths. On the other hand, because Tanh minimizes the path as it passes the function and does not make it zero, the path consisting of frequency can be created without missing value. Fig. 10 shows the result with ReLU activation function for 20,000 and 50,000 epochs. Comparing between Fig. 10(a) and Fig. 10(b), ReLU formed accumulated path group. However, ReLU generates clusters rather than generating paths even if epoch increases. Fig. 11 shows the results of two epoch processes with 20,000 and 50,000 epochs with Tanh activation function. About 20,000 epochs, ambiguous paths are created. Approximately more than 50,000 epochs, trained Generator creates a new path. Since the inputs of GAN model are diverse, the outputs seem to be slightly different but suitable paths.

### C. Unexplored Route

Fig. 12 shows two routes generated by GAN. Fig. 12-(a) shows typical output generated. Fig. 12-(b) is not trivial output comparing to output shown in Fig. 10-(a). GAN generated normal routes and it also generated additional routes. That is GAN creates unexplored, new route.

## VII. CONCLUSION

We developed a method to generate human mobility routes based on Generative Adversarial Networks (GANs) which are specialized in image generation. We successfully trained GAN model on individual mobility data, and the model could suggest novel routes in about 3 square kilometer of area, reflecting preference of specific object. In addition, we show the possibility to train any extensive trajectory pattern instead of massive geopositioning data by applying CNN layers in front of the model. The essence of our research is that GAN model can train on the continuous sequential data, such as trajectory pattern and can generate mobility route with additional information. The purpose of this model is

to provide unexplored routes which is based on an object's daily mobile data. By training an object's mobile data, we could generate mobility routes, which the users are likely to explore in their future. Therefore, generated mobility routes are fundamental basis to individual recommendation system. By combining our model with Location Based Service (LBS) and Recommender Systems will allow us to provide whole new service to users. Users will be recommended places such as personalized restaurants, shops, and other utilities within unexplored generated mobility routes.

While the better purpose in our experiment is to create mobile route on random location, the model's ability to generate the route would be bounded with the high randomness. Additionally, we did not interconnect between mobility data and human behavior factor. Further, the biased mobility data, which is adhere to specific location such as school or workplace, caused drawbacks in the model training, and the application based on geolocation system has inherent error. As we mentioned the combination of human behavior factor such as hobby, job, personality, etc., the research that combines with personal factor with our method would be an advanced method to generate mobility route more connected with mobility tendency, offering the personal customized solution. In future work, we aim to apply social media service which includes personal behavior features. It would be possible to identify the purpose of movement and specific time period. This can be expanded from individual to group with similar personality. As a result, it would be possible to create mobility route which is more customer-oriented.

## REFERENCES

- [1] D. Pfoser, C. S. Jensen, and Y. Theodoridis, "Novel approaches in query processing for moving object trajectories," in *Proceedings of the 26th International Conference on Very Large Data Bases*, ser. VLDB '00. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2000. ISBN 1-55860-715-3 pp. 395–406. [Online]. Available: <http://dl.acm.org/citation.cfm?id=645926.672019>
- [2] J. J.-C. Ying, W.-C. Lee, T.-C. Weng, and V. S. Tseng, "Semantic trajectory mining for location prediction," in *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ser. GIS '11. New York, NY, USA: ACM, 2011. doi: 10.1145/2093973.2093980. ISBN 978-1-4503-1031-4 pp. 34–43. [Online]. Available: <http://doi.acm.org/10.1145/2093973.2093980>
- [3] M. Gorawski and P. Jureczek, "Continuous pattern mining using the fcpgrowth algorithm in trajectory data warehouses," in *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2010, pp. 187–195. [Online]. Available: [https://doi.org/10.1007%2F978-3-642-13769-3\\_23](https://doi.org/10.1007%2F978-3-642-13769-3_23)
- [4] J. W. Lee, O. H. Paek, and K. H. Ryu, "Temporal moving pattern mining for location-based service," *Journal of Systems and Software*, vol. 73, no. 3, pp. 481–490, nov 2004. doi: 10.1016/j.jss.2003.09.021. [Online]. Available: <https://doi.org/10.1016%2Fj.jss.2003.09.021>
- [5] A. Monreale, F. Pinelli, R. Trasarti, and F. Giannotti, "Wherenext: a location predictor on trajectory pattern mining," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2009. doi: 10.1145/1557019.1557091 pp. 637–646.
- [6] H. Jeung, Q. Liu, H. T. Shen, and X. Zhou, "A hybrid prediction model for moving objects," in *2008 IEEE 24th International Conference on Data Engineering*. IEEE, apr 2008. doi: 10.1109/icde.2008.4497415. [Online]. Available: <https://doi.org/10.1109%2Ficde.2008.4497415>
- [7] M. Morzy, "Prediction of moving object location based on frequent trajectories," in *Computer and Information Sciences ISCIS 2006*. Springer Berlin Heidelberg, 2006, pp. 583–592. [Online]. Available: [https://doi.org/10.1007%2F11902140\\_62](https://doi.org/10.1007%2F11902140_62)

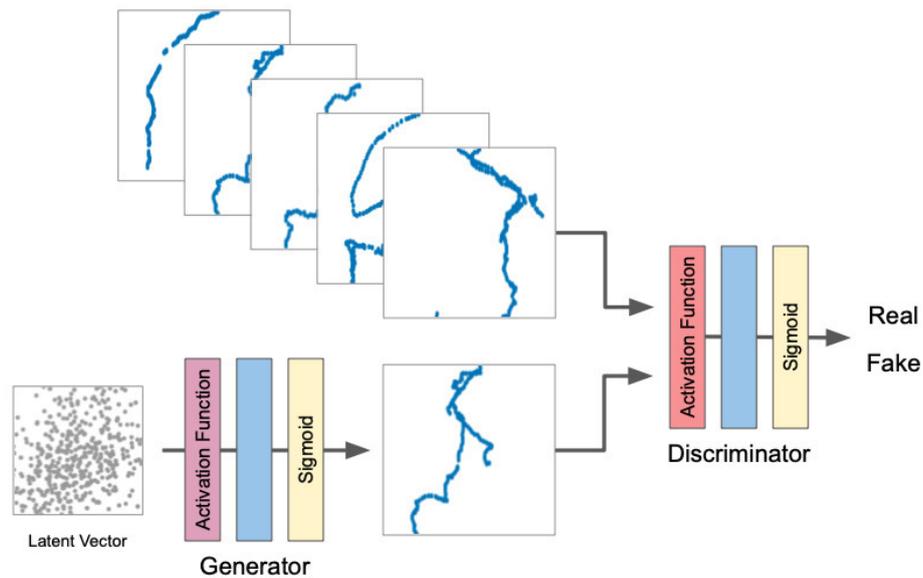


Fig. 6. Architecture for our GAN

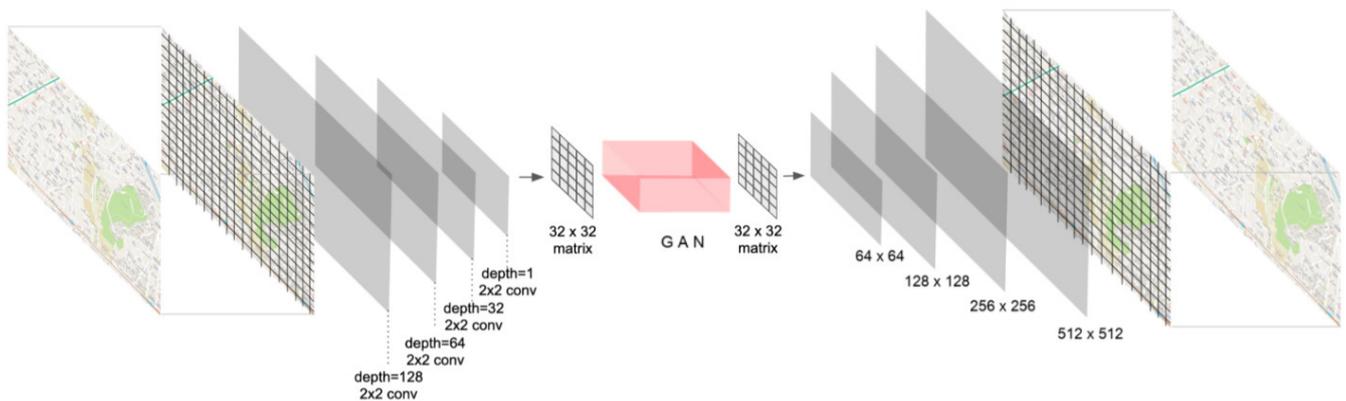


Fig. 7. Whole Architecture of our Model

- [8] —, “Mining frequent trajectories of moving objects for location prediction,” in *Machine Learning and Data Mining in Pattern Recognition*. Springer Berlin Heidelberg, 2007, pp. 667–680. [Online]. Available: [https://doi.org/10.1007%2F978-3-540-73499-4\\_50](https://doi.org/10.1007%2F978-3-540-73499-4_50)
- [9] F. Giannotti, M. Nanni, F. Pinelli, and D. Pedreschi, “Trajectory pattern mining,” in *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2007. doi: 10.1145/1281192.1281230 pp. 330–339.
- [10] V. T. H. Nhan and K. H. Ryu, “Future location prediction of moving objects based on movement rules,” in *Intelligent Control and Automation*. Springer Berlin Heidelberg, 2006, pp. 875–881. [Online]. Available: [https://doi.org/10.1007%2F11816492\\_112](https://doi.org/10.1007%2F11816492_112)
- [11] H. Y. Song and D. Y. Choi, “Defining measures for location visiting preference,” *Procedia Computer Science*, vol. 63, pp. 142–147, 2015. doi: 10.1016/j.procs.2015.08.324. [Online]. Available: <https://doi.org/10.1016%2Fj.procs.2015.08.324>
- [12] A. Sudo, T. Kashiya, T. Yabe, H. Kanasugi, X. Song, T. Higuchi, S. Nakano, M. Saito, and Y. Sekimoto, “Particle filter for real-time human mobility prediction following unprecedented disaster,” in *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ser. SIGSPACIAL '16. New York, NY, USA: ACM, 2016. doi: 10.1145/2996913.2997000. ISBN 978-1-4503-4589-7 pp. 5:1–5:10. [Online]. Available: <http://doi.acm.org/10.1145/2996913.2997000>
- [13] M. Baratchi, N. Meratnia, P. J. M. Havinga, A. K. Skidmore, and B. A. K. G. Toxopeus, “A hierarchical hidden semi-markov model for modeling mobility data,” in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp 14 Adjunct*. ACM Press, 2014. doi: 10.1145/2632048.2636068. [Online]. Available: <https://doi.org/10.1145%2F2632048.2636068>
- [14] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” *arXiv preprint arXiv:1511.06434v2*, 2016.
- [15] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, “StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Jun 2018. doi: 10.1109/cvpr.2018.00916. [Online]. Available: <https://doi.org/10.1109%2Fcvpr.2018.00916>
- [16] M. Molano-Mazon, A. Onken, E. Piasini, and S. Panzeri, “Synthesizing

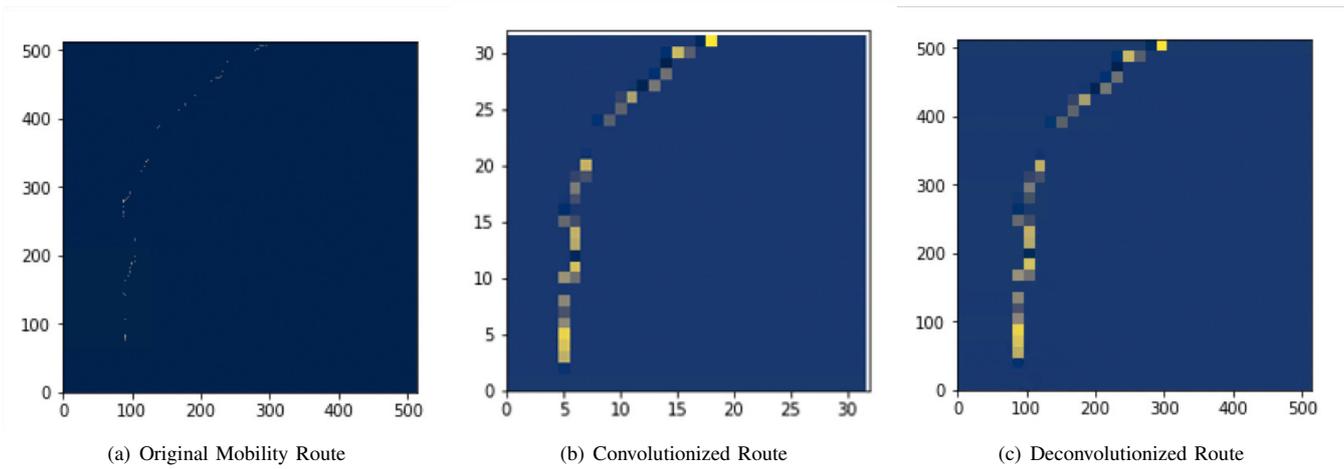


Fig. 8. Result of Convolution and Deconvolution

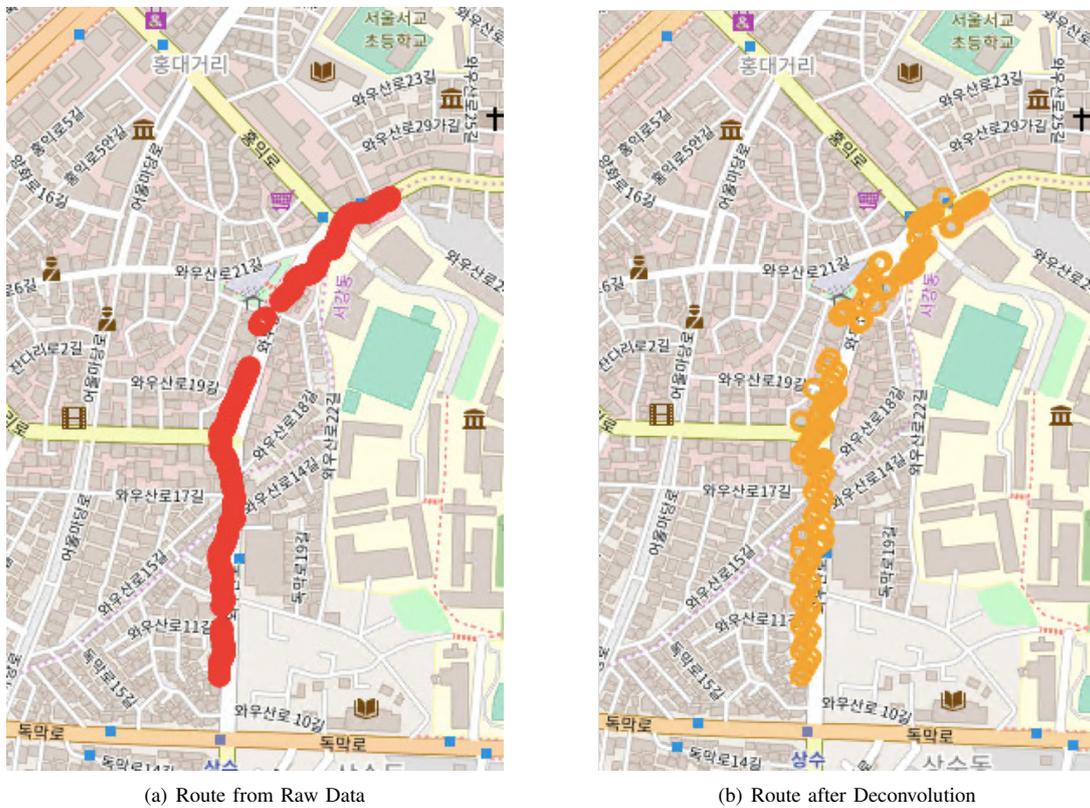


Fig. 9. Mapped Result from Raw Mobility Data versus after Deconvolution

realistic neural population activity patterns using generative adversarial networks,” *arXiv preprint arXiv:1803.00338*, 2018.

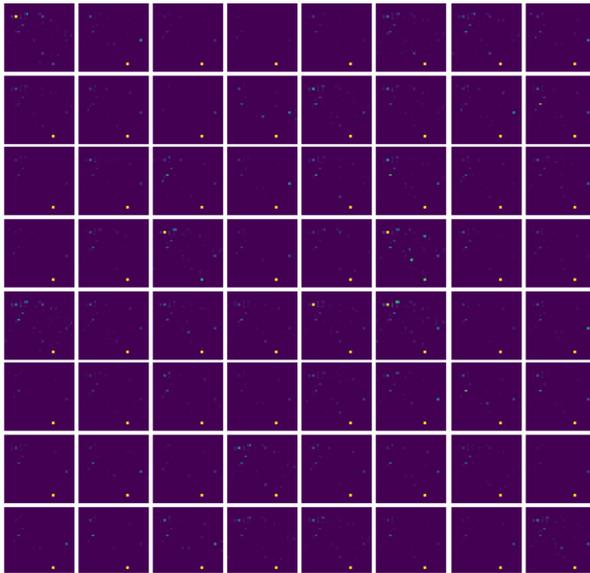
[17] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi, “Social gan: Socially acceptable trajectories with generative adversarial networks,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Jun 2018. doi: 10.1109/cvpr.2018.00240. [Online]. Available: <https://doi.org/10.1109%2Fcvpr.2018.00240>

[18] M. Alzantot, S. Chakraborty, and M. Srivastava, “Sensegen: A deep learning architecture for synthetic sensor data generation,” in *2017 IEEE International Conference on Pervasive Computing*

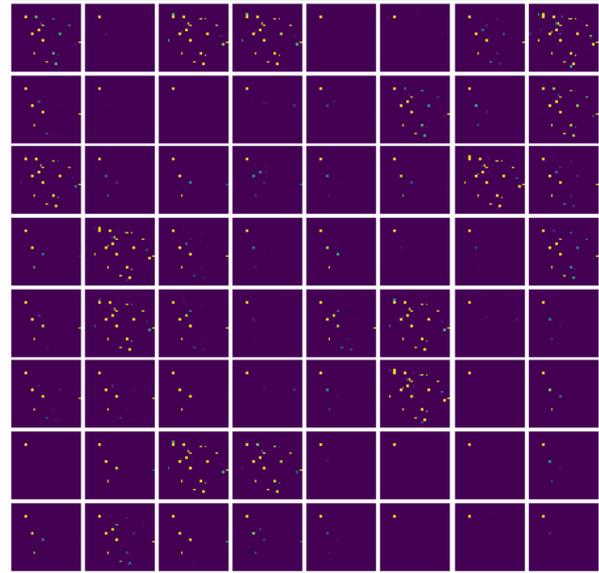
and Communications Workshops (PerCom Workshops). IEEE, Mar 2017. doi: 10.1109/percomw.2017.7917555. [Online]. Available: <https://doi.org/10.1109%2Fpercomw.2017.7917555>

[19] S. Na, L. Xumin, and G. Yong, “Research on k-means clustering algorithm: An improved k-means clustering algorithm,” in *2010 Third International Symposium on intelligent information technology and security informatics*. IEEE, 2010. doi: 10.1109/IITSI.2010.74 pp. 63–67.

[20] P. Reinecke, T. Krauss, and K. Wolter, “Hyperstar: Phase-type fitting made easy,” in *2012 Ninth International Conference on Quantitative*

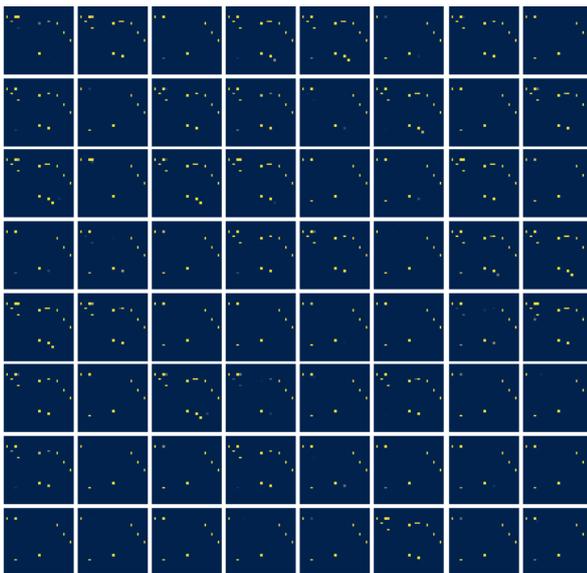


(a) 20,000 epochs with ReLU

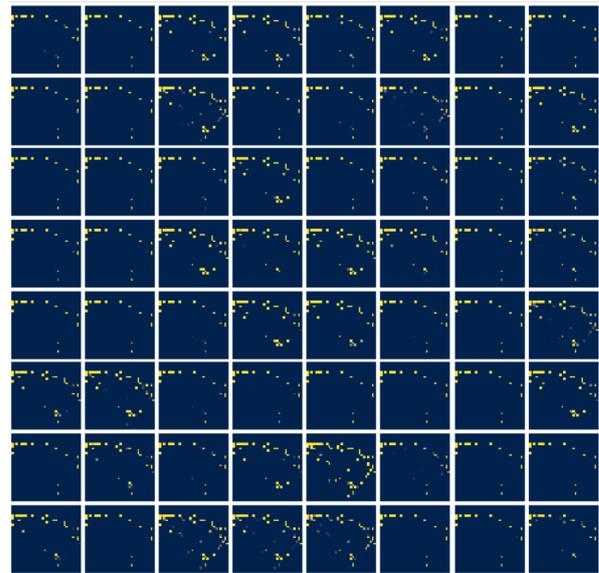


(b) 50,000 epochs with ReLU

Fig. 10. Result with ReLU activation function



(a) 20,000 epochs with Tanh



(b) 50,000 epochs with Tanh

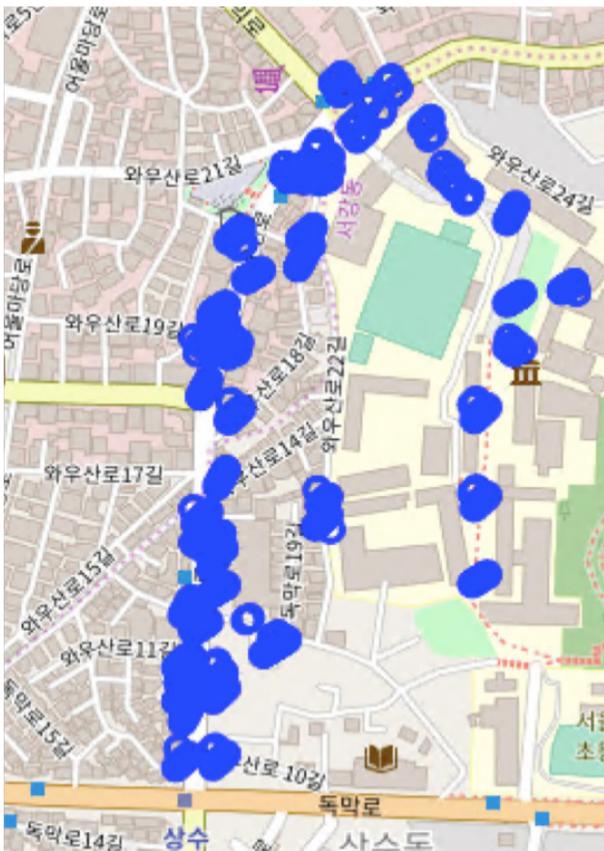
Fig. 11. Result with Tanh activation function

*Evaluation of Systems*. IEEE, Sep 2012. doi: 10.1109/qest.2012.29. [Online]. Available: <https://doi.org/10.1109/qest.2012.29>

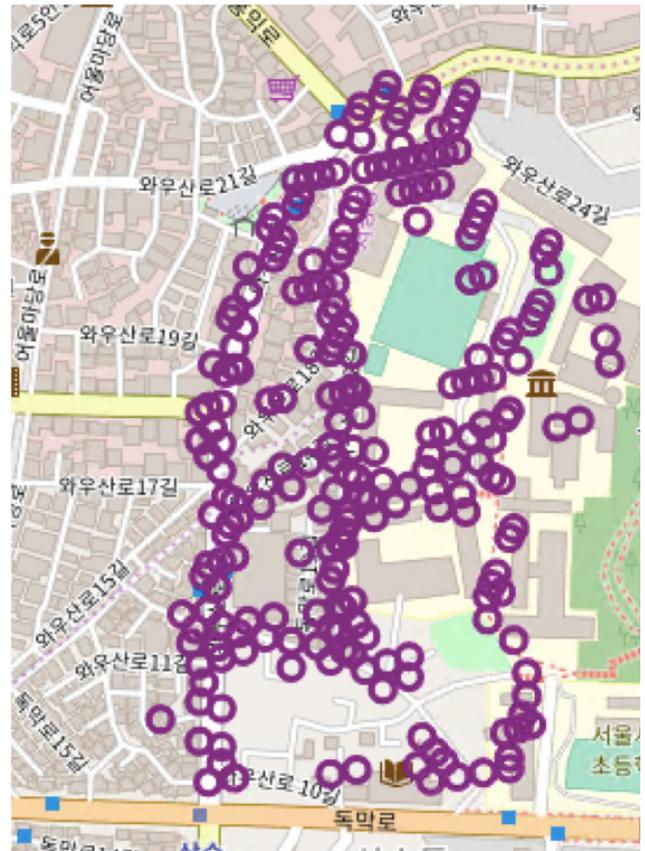
- [21] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds.

Curran Associates, Inc., 2014, pp. 2672–2680. [Online]. Available: <http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf>

- [22] W. Fedus, M. Rosca, B. Lakshminarayanan, A. M. Dai, S. Mohamed, and I. Goodfellow, "Many paths to equilibrium: Gans do not need to decrease adivergence at every step," 10 2017.



(a) General Generated Route



(b) Generated Route with Additional Information

Fig. 12. Route Generated by GAN



# A Deep Learning and Multimodal Ambient Sensing Framework for Human Activity Recognition

Ali Yachir, Abdenour Amamra, Badis Djamaa, Ali Zerrouki and Ahmed khierEddine Amour  
*Military Polytechnic School*  
*PO BOX 17, Bordj-El-Bahri, 16111, Algiers, Algeria*  
*Email: {ali.yachir, amamra.abdenour, badis.djamaa}@gmail.com*

**Abstract**— Human Activity Recognition (HAR) is an important area of research in ambient intelligence for various contexts such as ambient-assisted living. The existing HAR approaches are mostly based either on vision, mobile or wearable sensors. In this paper, we propose a hybrid approach for HAR by combining three types of sensing technologies, namely: smartphone accelerometer, RGB cameras and ambient sensors. Acceleration and video streams are analyzed using multiclass Support Vector Machine (SVM) and Convolutional Neural Networks, respectively. Such an analysis is improved with the ambient sensing data to assign semantics to human activities using description logic rules. For integration, we design and implement a Framework to address human activity recognition pipeline from the data collection phase until activity recognition and visualization. The various use cases and performance evaluations of the proposed approach show clearly its utility and efficiency in several everyday scenarios.

## I. INTRODUCTION

The combination of the Ambient Intelligence and the Internet of Things [1] aims at building smart environments by integrating a variety of interconnected devices such as camera, smartphone, smart watch and actuator. Such a sensing and actuating technology, has allowed to the analysis of human daily activities to become easier and straightforward. Particularly, in smartly controlled environments such as smart home, HAR can be envisioned for several potential applications and different contexts including security, healthcare, ambient assisted living and behavior analysis. For instance, many HAR systems surveyed in [2, 3], where the authors focus on different activities (walking, running, cooking, exercising, etc.) in different application domains.

In practice, there are diverse ways of using sensors for human activity recognition in a smart environment. Hence, the existing approaches can be divided into two main categories, namely: vision-based and sensors-based approaches. In the former approaches, the primitive actions of an activity are detected by analyzing the images transmitted by an RGB camera. Such an analysis can exploit computer vision techniques to recognize patterns. Whereas the latter approaches (sensor-based) use sensors that are either worn by a person or placed on everyday objects. Wearable sensors can be placed on clothing, in a pocket, or stuck directly to the body (wrist, hip or torso) to provide valuable information about an individual's degree of functional ability and lifestyle [4]. Indeed, sensors' position

should be well chosen in order to ensure their usability while offering a maximum comfort to the user. In addition, sensors can be placed seamlessly on ordinary objects to detect and control the environment. They can also be of different types such as: contact detectors to give the state (close/open) of doors and cabinets, pressure mats to indicate the position of the person in the room or to detect if a person is sitting on a sofa or laying on a bed, RFID tags to give the location of objects, etc. According to a recent study [5], the RGB cameras have lower popularity when compared to depth sensors and wearable devices in HAR research.

In order to implement a HAR system, the data collected and transmitted by various cameras and sensors disseminated in the smart environment can be analyzed using several techniques in either vision or sensors-based approaches. Regarding vision-based approaches, a survey of action recognition approaches based on Space-Time Interest Points (STIP) was proposed in [6]. Most recent approaches are based on Convolutional Neural Networks (CNNs) including Deep Convolutional Networks (ConvNets) [7] and TwoStream [8]. These deep learning methods aim to learn automatically the semantic representation of raw videos by using a deep neural network in a discriminatory manner from a large number of tagged data. For analyzing real-time videos, Recurrent Neural Networks (RNN) among which there are Long Short-Term Memory (LSTM) units have been proposed. LSTM networks have proved their effectiveness in several areas such as: images and videos subtitling [9] and temporal information of movements and videos streams. Regarding sensors-based approaches, a deep ConvNet was also used in [10] to perform HAR using smartphone sensors by exploiting the inherent characteristics of activities. In [11], acceleration streamed by a smartphone are analyzed with K-Nearest Neighbors (KNN) for recognizing several types of activities (walking, climbing, sitting, standing and falling down). In [12], data from inertial and pressure sensors placed on the trunk of a patient are used to recognize activities such as walking, sleeping and climbing stairs. In [13], Hidden Markov Model (HMM) is used to classify complex actions such as running, walking or laying, using the accelerometer data of a wristwatch. In [14], simple and complex activities such as cleaning, hand washing, and plant watering are recognized using fixed window lengths with an overlapping halved window. In [15], human activity recognition is analyzed

through the segmentation of the multidimensional time series of acceleration data based on a specific multiple regression model. In [16], a digital low-pass filter is designed to recognize certain types of human physical activities using acceleration data. In [17], the selection and placement of wearable sensors is investigated for classifying sixteen activities of daily living for six healthy subjects.

The aforementioned discussed works show that most of the proposed approaches recognize simple human activities such as laying, sitting, and standing. Moreover, these approaches focus on the data received from either cameras or other sensors without a real combination of the different modalities that can become unavailable due to their temporary or permanent disappearance, and should therefore, be replaced to ensure HAR continuity. Furthermore, contextual information such as localization, acceleration and object state provided by mobile or wearable sensors combined with machine learning methods offer a higher accuracy and diversity for recognizing complex human activities (watching TV, cooking, exercising, etc.).

We propose a hybrid approach for HAR in an ambient environment by combining three types of sensing technologies, namely: smartphone accelerometer, RGB cameras and ambient sensors. First, real time accelerations and video streams are analyzed separately using machine learning algorithms to detect and recognize simple human activities or postures. Video streams are used by default for indoor spaces, but they are replaced by smartphone accelerometer data in the case of inaccessible cameras. Switching between these two modes can considerably increase the reliability of the designed HAR system. Second, additional information is extracted from the available activated ambient sensors to assign semantics to human activity using Description Logic (DL) rules. Finally, the three types of the provided information are combined inside a HAR framework using supervised machine learning algorithms in order to recognize and visualize more complex activities.

The remainder of the paper is organized as follows. In Section 2, we describe our acceleration-based activity recognition method. In Section 3, the video-based activity recognition method is explained. In Section 4, we present the hybrid approach and the designed framework for complex activity recognition. In section 5, we conduct several validation scenarios for the recognition of everyday activities. The paper is concluded with section 6, and potential future works are announced.

## II. ACCELERATION BASED ACTIVITY RECOGNITION

The acceleration data along the three axes ( $x$ ,  $y$ , and  $z$ ) is collected from a smartphone worn on the waistband of the user's pelvis. This data collection operation is performed by an Android application with a sampling rate of 50 Hz, i.e. the data is divided into a window of 50 records per second. We distinguished six (6) classes of elementary actions or postures namely: sitting, standing, running, walking, walking upstairs and walking downstairs. We collected 500 records for each class. For a better distinction between the different classes, we chose, as an input to our machine learning model, a vector of 30 characteristics such as average, variance, and min-max with respect to  $x$ ,  $y$  and  $z$ ; resulting average of the acceleration; AR-

coefficient; Angle Tilt; and Signal Magnitude Area (SMA) [18]. We opted for this choice after performing several tests by combining these different characteristics. For each combination, we calculate classification success rate. These characteristics ensure a high degree of independence between the different classes and minimize the correlation between them. When constructing the learning model, we tested six (06) learning algorithms, which were Naïve Bayes, SVM with linear kernel, SVM with rbf kernel, nonlinear SVM, k-Nearest Neighbors (kNN) and MultiLayer Perceptron (MLP) with a single hidden layer. The success rates obtained from these algorithms are shown in the **Table I**. The latter shows that SVM with linear kernel gives the best performance (93%).

TABLE I. SUCCESS RATE OF THE TESTED ALGORITHMS

Approach	Success rate
Naïve Bayes	92%
SVM with linear kernel	<b>93%</b>
SVM with rbf kernel	90%
Nonlinear SVM	92%
k-Nearest Neighbors (kNN)	87%
MultiLayer Perceptron (MLP)	87%

## III. VISION BASED ACTIVITY RECOGNITION

### A. Dataset construction

The dataset is constructed using two different sources: Multiple Pose Human Body Database (LSP / MPII-MPHB) [19] and other data that we gathered from Google Image search engine. First, the LSP / MPII-MPHB contains 26675 images and 29732 human bodies that are divided into six (06) action categories: curving, knee, laying, occlusion, sitting and standing. For each image, we detect the persons using the Single Shot MultiBox Detector (SSD) method [20], which is a unified Framework for detecting objects with a single neural network. We focus only on three main postures: standing, sitting and laying. Then, we used Google Image search facility to retrieve all possible images for these three main postures.

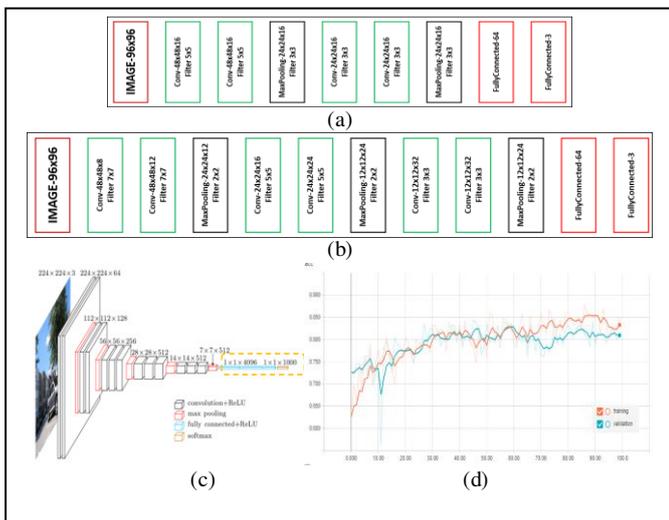
After building the dataset, we obtain a considerable number of images for each action or posture category. The images represent the inputs of the learning model while the actions represent the classes respective to the outputs. Thus, in order to standardize and reduce the size of the training data and to speed up their processing, we perform a pre-processing on all the input images. First, each image is stored with a reduced size of 224x224 pixels and three channels representing the values of the three colors: red, green and blue. Then, the images are normalized and scaled at loading time by a centering and reduction transformation in the interval [-1,1]. Finally, we split the images into three folders: sitting, standing and laying down to assign them to the different classes. In order to feed our learning model, we transformed the set of images into a four-dimensional (height 224 pixels, width 224 pixels and the three-color channels) tensor and the class labels into a one-dimensional vector.

### B. Learning model

We proposed two architectures of convolutional neural networks, where each is composed of several layers of different

types. We also tested a learning method based on the transfer of parameters. These relatively deep architectures have a structure inspired by that of the VGGNet network [20]. The difference between them lies in the depth and the number of hyperparameters. Both architectures are composed of large blocks; where the initial blocks are constituted of two convolutional layers followed by a pooling layer, and the last ones are composed of only dense layers. The depth of the convolutional layers increases from one block to another, although the spatial size of the filters decreases. The choice of convolutional layers comes from the fact that they are the most adapted to image recognition tasks as they consider the multidimensional aspect of images. Moreover, each neuron in these layers is connected to just a small set of neurons in the preceding layer, the number of learnable parameters is therefore smaller due to the parameter sharing property of convolution. The  $z$  outputs of each convolutional layer are filtered by the Rectified Linear Unit (ReLU) activation function.

Fig. 1. Our activity recognition architectures, (a), Architecture 1, (b) Architecture 2, (c) Architecture 3, (d) Results of our activity recognition CNN



A maxpooling layer follows each pair of convolutional layers. These layers lead to the reduction of the dimensions of the feature maps by applying the *max* function on a window of neighboring pixels at a given region of the image. Therefore, the maxpooling reduces the intraclass variance by discarding the unnecessary information. At the last level of the network, a dense layer is considered in order to gather all the features detected throughout the network. The output layer is also a dense layer whose number of neurons is equal to the number of classes (in our case, the number of classes is set to 3). The  $z$  results are passed through a *softmax* function in order to be squashed into the interval  $[0,1]$  leading to a probability distribution over the classes. The reason why the stacked architecture is preferable is that the first layers detect low-level features (such as edges and simple shapes) whereas the deeper ones detect high-level features (such as complex shapes and objects).

In what follows, we present the three architectures that we proposed and tested in the light of this contribution.

### C. Architecture 1

The first block of the network consists of two convolutional layers having each a depth of 16 feature maps and a filter size of  $5 \times 5$ , with stride 1 horizontally and vertically, and zero padding on all four edges. These layers are followed by a maxpooling layer with a filter of size  $2 \times 2$  and a step of 2, meaning that we reduce the dimensions of the input by a half. The second block is identical to the first one except that the depth of each convolutional layer is 32 instead of 16 and the size of the convolution filter is  $3 \times 3$ .

The third block, is a dense layer of 64 neurons connected to all the outputs from the previous maxpooling layer. The output layer is of size 3, where each neuron corresponds to a class. **Figure 1 (a)** illustrates Architecture 1. The green rectangles represent the convolutional layers, the blacks represent the pooling layers, and the reds represent the dense layers.

### D. Architecture 2

This architecture is similar to the first one but it is composed of four main blocks. The first block encompasses two convolutional layers with a filter of  $7 \times 7$  and a depth of 8 and 12, respectively. The second contains similar layers whose filter size is  $5 \times 5$  and depths is 16 and 24, respectively. The two layers of the last block have a  $3 \times 3$  filter and a depth of 32 each. Similarly, a pooling layer follows each pair of convolutional layers in all the blocks. Identically to the previous architecture, this one contains a dense layer of 64 neurons in the last block (see **Figure 1 (b)**).

### E. Architecture 3

The first two architectures, that we proposed, are prone to overfitting because the training dataset is small compared to the size of the network. In order to avoid such a drawback, we used a Transfer Learning approach and we augmented our dataset by applying several image transformations. We initialized the new model with the weights of VGG16 network trained on our dataset. During the training, we maintained a fixed number of layers (the first convolutional layers) and we optimized the parameters of the latter ones. Our purpose is to reduce optimization space dimension, whilst reusing the first convolutional blocks. The latter are most likely to remain similar regardless of the problem at hand, and the model must be able to fit specific top-level layers. We adapted the dense layer with our classes of actions. **Figure 1 (c)** shows the modifications made to the VGG16 architecture for transfer learning. Our best results for activity recognition were obtained with the third architecture (**Figure 1 (c)**).

### F. Cost function and the optimizer

The neural networks are general functions estimators. Nevertheless, their estimation is never perfect as there is always a discrepancy between the output of the network and the ground truth values. Therefore, we define a cost function to measure the error and we adapt an optimization algorithm to minimize its value. Given the transformation of this problem into a classification task, the most appropriate interclass error measure is the cross-entropy function [21]. The learning is achieved by minimizing the norm of the cross-entropy with gradient descent algorithm. Such an algorithm

consists in changing the weights of the network by a factor of a fixed learning rate. We chose the *Adadelta* [22] optimizer as it does not depend.

#### G. Training platform and implementation

One of the major problems in deep learning is the significant requirement for computation resources during the training. In order to alleviate this problem, we trained our algorithm on a machine equipped with a Nvidia GTX 860 GPU and 12Gb of RAM. For the purpose of robustness and versatility, we implemented our CNN in Python3 using TensorFlow library. This library runs a low-level C++ routine, which are able to perform massively parallel computations using all the available processing power and benefiting from the existing hardware vectorization of the GPU. Training and validation progress can be seen in **Figure 1 (d)**

#### IV. HYBRID APPROACH FOR HUMAN ACTIVITY RECOGNITION

The two previously presented models, based on accelerometer and camera, are able to recognize elementary actions or postures of the person (e.g., standing, sitting and laying) while an activity is defined as a task of daily life that the person performs over a given time interval. Thus, other contextual information can complement the information of the posture to deduce the effective activity. The localization of the subject, for example, is an essential attribute, due to the fact that most activities are carried out in a specific room. For instance, the "cooking" activity takes place in the kitchen while "Watching TV" is more likely to appear in the living room. In addition, the posture of the person is a influential factor to recognize an activity. We cannot imagine, for example, a person sleeping in a "standing" posture. It should also be noted that the actions related to a particular activity can activate or deactivate a number of ambient sensors. In order to represent all these causal relationships, we used DL rules to infer five main activities namely: watching TV, sleeping, preparing a meal, communicating (talking on the phone) and working with a laptop or PC. For example, the activity "Sleeping" can be inferred using the two DL rules shown in **Table II**.

TABLE II. EXAMPLE OF DL RULES FOR "SLEEPING" ACTIVITY

$\text{Sleeping} \equiv \text{Posture}(\text{laying\_down}) \cap$ $\text{Location}(\text{bedroom}) \cap$ $\text{AmbientSensors}(\text{Bed}).\text{HasStatus}(\text{On}) \cap$ $\text{AmbientSensors}(\text{Light}).\text{HasStatus}(\text{Off})$
$\text{Sleeping} \equiv \text{Posture}(\text{laying\_down}) \cap$ $\text{Location}(\text{living\_room}) \cap$ $\text{AmbientSensors}(\text{Sofa}).\text{HasStatus}(\text{On}) \cap$ $\text{AmbientSensors}(\text{Light}).\text{HasStatus}(\text{Off}) \cap$ $\text{AmbientSensors}(\text{TV}).\text{HasStatus}(\text{Off})$

Since we can have several rules for the same activity, we transformed these rules to a description vector containing posture, location and the state of the other sensors as shown in **Table III**. Hence, all the data contained in description vectors summarize a temporal window. Such vectors are used as an input for the SVM algorithm with linear kernel and their

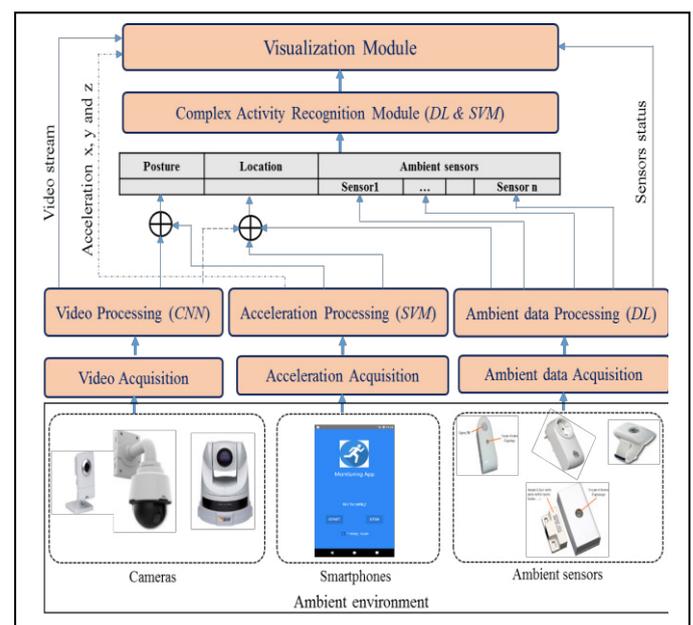
corresponding activities, in DL rules, represent the output. The supervised learning model is constructed by generating a set of input vectors with the identifier of the corresponding class or activity. These vectors contain the different possible combinations of values regarding the considered characteristics. For example, in "Watching TV" activity, the generated posture can be "sitting" or "laying down". If a feature vector does not match any of the existing classes, the person's activity is considered as "Unknown".

TABLE III. STRUCTURE OF THE INPUT LEARNING VECTOR

Posture	Location	Ambient sensors			
		Sensor1	...	Sensor n	

We developed a framework for human activity recognition in ambient environment. As shown in **Figure 2**, the general architecture of the proposed framework allows data acquisition and processing as well as activity recognition and visualization. First, heterogeneous data are collected using different sensing modalities, namely: cameras for real-time video streams, the various sensors of the Smartphone (accelerometer, gyroscope, etc.) as well as the others ambient environment sensors (pressure, temperature, humidity, light, movement, etc.). Then, collected data are processed separately using CNN, SVM and DL respectively for video, acceleration and ambient sensors. We should notice that the posture is recognized, by default, using video stream, but automatic switching to acceleration is performed in case of cameras unavailability to ensure service continuity. The location can be also deduced from cameras and smartphones' position or provided by other sensors. Finally, a description vector is constructed from all the previous processed data and used as input of the machine learning model to recognize more complex activities. The recognized activity along with the collected data from camera, smartphone and ambient sensors are transmitted to the visualization module for display and monitoring in real time.

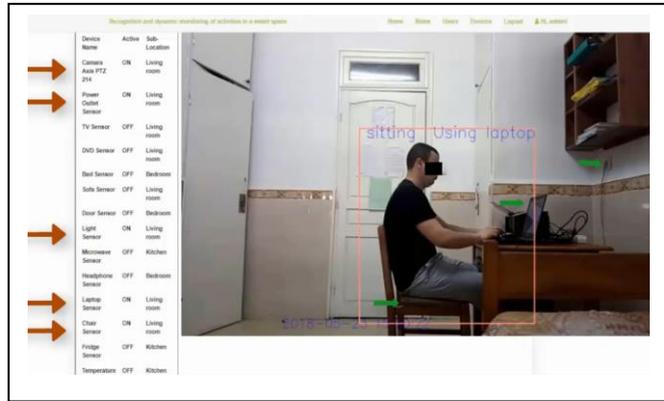
Fig. 2. Global architecture of the designed HAR Framework



V. IMPLEMENTATION AND USE CASES SCENARIOS

The proposed Framework has been applied in several situations for human activity recognition. For example, **Figure 3** shows a person using a personal computer. This activity is recognized with the combination of several information: the posture is "sitting", the location is "office", the pressure sensor on the chair is activated "ON" and the laptop is "ON".

Fig. 3. Example of recognized activity "Working with laptop"



Regarding the performance evaluation, we achieved a success rate of 97% after several classification tests of the proposed HAR approach of the designed framework. As shown in the **Table IV**, our approach gives a better result considering five classes of activities and several multi-source data (ambient sensors, cameras and acceleration). The result of 98.2% obtained in [12] is due to the fact that the authors used the acceleration by considering only three classes of activities which further minimizes the ambiguity between classes.

TABLE IV. COMPARISON OF SUCCESS RATES OF ACTIVITY RECOGNITION

Approach	Devices	Classifier	Success rate
[12]	Inertial and barometric sensor	KNN	98.2%
[13]	Accelerometer of a wristwatch	HMM, CRF	90.4%
[15]	Accelerometer on Chest, Thigh and Ankle	Multiple regression model	90.3%
[16]	Accelerometer on Hand and pocket	digital low-pass filter	91.15%
[17]	Pressure sensor	KNN	89.08%
Our approach	Cameras, smartphone, ambient sensors	CNN, DL and SVM with linear kernel	97%

VI. CONCLUSION AND FUTURE WORK

We proposed a hybrid solution for human activity recognition using smartphone inertial sensors (accelerometers), RGB cameras and ambient sensors (pressure, localization, etc.). Acceleration data and videos were analyzed using machine learning algorithms, SVM and CNN in order to detect the current potential posture of the person. Such an analysis is augmented with ambient sensor data to assign semantics to the human activity based on description logic rules. A HAR framework is also designed to build the whole pipeline from data collection until activity recognition and visualization. We

are currently working to use our approach in the context of ambient-assisted living to assist elderly or dependent persons to improve their quality-of-life. Future works will include further experimentation and combination of other techniques such as automatic image captioning using deep learning.

REFERENCES

- [1] P. Suresh, J. V. Daniel, V. Parthasarathy, and R. H. Aswathy, "A state of the art review on the Internet of Things (IoT) history, technology and fields of deployment," in Proc. IEEE Int. Conf. Sci., Eng. Manage. Res. (ICSEMR'14), Chennai, India, Nov. 2014, pp. 1-8.
- [2] O. Lara, M. Labrador, "A survey on human activity recognition using wearable sensors", IEEE Commun. Surv. Tutor. 1 (2012) 1-18.
- [3] H.F. Nweke, Y.W. Teh, M.A. Al-garadi, and U.R. Alo, "Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges" Expert Systems With Applications 105 (2018).
- [4] M.M. Hassan, M. Zia Uddin, A. Mohamed, and A. Almgren: "A robust human activity recognition system using smartphone sensors and deep learning" in Future Generation Computer Systems 81 (2018) 307-313.
- [5] O.C. Ann, L.B. Theng, "Human activity recognition: A review" in 2014 IEEE International Conference on Control System, Computing and Engineering.
- [6] D.D. Dawn and S.H. Shaikh, "A comprehensive survey of human action recognition with spatio-temporal interest point (stip) detector," The Visual Computer, vol.32, no. 3, pp. 289-306, 2016.
- [7] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Largescale video classification with convolutional neural networks," in Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2014, pp. 1725-1732.
- [8] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," in Advances in neural information processing systems, 2014, pp. 568-576.
- [9] S. Yousfi, "Embedded Arabic text detection and recognition in videos," Ph.D. dissertation, Lyon University, 2016.
- [10] C.A. Ronao, S.-B. Cho "Human activity recognition with smartphone sensors using deep learning neural networks" Expert Systems With Applications 59 (2016).
- [11] Y. Kwon, K. Kang, and C. Bae, "Unsupervised learning for human activity recognition using smartphone sensors," Expert Systems with Applications, vol.41, no. 14, pp. 6067-6074, 2014.
- [12] F. Massé, R. R. Gonzenbach, A. Arami, A. Paraschiv-Ionescu, A. R. Luft, and K. Aminian, "Improving activity recognition using a wearable barometric pressure sensor in mobility-impaired stroke patients," Journal of neuroengineering and rehabilitation, vol. 12, no. 1, p. 72, 2015.
- [13] E. Garcia-Ceja, R. F. Brena, J. C. Carrasco-Jimenez, and L. Garrido, "Long-term activity recognition from wristwatch accelerometer data," Sensors, vol. 14, no. 12, pp. 22 500-22 524, 2014.
- [14] S. Dermbach, B. Das, N. C. Krishnan, B. L. Thomas, and D. J. Cook, "Simple and complex activity recognition through smart phones," in Intelligent Environments (IE), 2012, pp. 214-221.
- [15] F. Chamroukhi, S. Mohammed, D. Trabelsi, L. Oukhellou, and Y. Amirat, "Joint segmentation of multivariate time series with hidden process regression for human activity recognition," Neurocomputing, vol. 120, pp. 633-644, 2013.
- [16] A. Bayat, M. Pomplun, and D. A. Tran, "A study on human activity recognition using accelerometer data from smartphones," Procedia Computer Science, vol. 34, pp. 450-457, 2014.
- [17] A. Moncada-Torres, K. Leuenberger, R. Gonzenbach, A. Luft, and R. Gassert, "Activity classification based on inertial and barometric pressure sensors at different anatomical locations," Physiological measurement, vol. 35, no. 7, 2014.
- [18] A. M. Khan, Y.-K. Lee, and T.-S. Kim, "Accelerometer signal-based human activity recognition using augmented autoregressive model coefficients and artificial neural nets," in Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE. IEEE, 2008, pp 5172-5175
- [19] Y. Cai and X. Tan, "Weakly supervised human body detection under arbitrary poses," in Image Processing (ICIP), 2016, pp. 599-603.
- [20] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd : Single shot multibox detector," in European conference on computer vision. Springer, 2016, pp. 21-37.
- [21] I. Goodfellow, Y. Bengio, and A. Courville, "Deep Learning". MIT Press, 2016.
- [22] S. Ruder, "An overview of gradient descent optimization algorithms," CoRR, vol. abs/1609.04747, 2016.



# Machine-learning at the service of plastic surgery: a case study evaluating facial attractiveness and emotions using R language

Lubomír Štěpánek

Department of Biomedical Informatics  
Faculty of Biomedical Engineering  
Czech Technical University in Prague  
Kladno, Czech Republic  
lubomir.stepanek@fbmi.cvut.cz

Jan Měšťák

Department of Plastic Surgery  
First Faculty of Medicine  
Charles University and  
Na Bulovce Hospital  
Prague, Czech Republic  
jan.mestak@lf1.cuni.cz

Pavel Kasal

Department of Biomedical Informatics  
Faculty of Biomedical Engineering  
Czech Technical University in Prague  
Kladno, Czech Republic  
pavel.kasal@fbmi.cvut.cz

**Abstract**—Since the plastic surgery should consider that facial impression is always dependent on current facial emotion, it came to be verified how precise classification of facial images into sets of defined facial emotions is.

Multivariate regression was performed using R language to identify indicators increasing facial attractiveness after undergoing rhinoplasty. Bayesian naive classifiers, decision trees (CART) and neural networks, respectively, were applied to assign a landmarked facial image data into one of the facial emotions, based on Ekman-Friesen FACS scale.

Enlargement of nasolabial and nasofrontal angle within rhinoplasty significantly predicts facial attractiveness increasing ( $p < 0.05$ ). Decision trees showed the geometry of a mouth, then eyebrows and finally eyes affect in this descending order an impact on classified emotion. Neural networks proved the highest accuracy of the classification.

Performed machine-learning analyses pointed out which geometric facial features increase facial attractiveness the most and should be consequently treated by plastic surgeries.

## I. INTRODUCTION

**F**ACIAL attractiveness was evaluated far earlier than origins of plastic facial surgery are dated. Whereas origins of plastic facial surgery are related to the First World War (1914–1918), human facial attractiveness received attention from ancient philosophers Polykleitos and Aristotle (4–3 century BC) [1]. Ancient classical rules were defined only subjectively and were strongly limited to the Caucasian race facial appearance as well as based only on viewing of beauty by a naked eye [1].

During the period of Renaissance, Leonardo Da Vinci modernized the classical rules of facial attractiveness viewing and refined them into so-called Neoclassical facial canons, based on the ancient principles. Neoclassical facial canons served mostly for contemporary artists and consisted of nine simple mathematical rules in the terms of *a subtraction or proportion of two linear facial distances should be equaled to a fixed constant*, e. g. “a maximal nose width should be one quarter of overall width of a face” etc.

The rules of the Neoclassical facial canons are still applied – if technically possible – to current plastic facial surgery procedures. An idea that some proportions of selected two different facial distances should be approximately equal to the golden ratio  $\left(\frac{\sqrt{5}-1}{2}\right)$ , is typical for the Neoclassical canons [2]. Similarly, flawlessly or nearly-perfectly axially-symmetric faces [3] and faces very similar to the *mean face of a population*, i. e. morphed facial shapes based on graphical averaging all facial control points of a bunch of faces using the given population, are generally considered as attractive ones [4]. Signs of human faces called *neoteny* (juvenilization), i. e. relative large eye and small mouth sizes, are also associated with a higher level of attractiveness [5]. Finally, sexual dimorphism plays a role in the perception of human facial attractiveness – both male faces with prevailing masculine facial geometry and female faces with prevailing feminine facial geometry are seems to be evaluated as more attractive [6].

All the mentioned rules, the Neoclassical Canon inclusively, are still commonly applied in nowadays plastic facial surgery procedures, including rhinoplasty, and – what is more – they are the principal (or even only) ways of how operational strategies are planned. However, saying this, data-driven approach to techniques covered by plastic facial surgery is the one whose time has to come [7].

Current demands of patients undergoing plastic facial surgeries include wishes handling with a not only improvement of “static” facial features such as corrections of nasal size or shape (rhinoplasty), but also changes for the better of the “dynamic” facial expression, e. g. surgical changes of mouth in order to make a smile more facial-appealing and to increase the facial attractiveness level for only moments when a patient’s face is smiling [7]. This is why movements of facial muscles during emotion expression and their connection to the facial impressions should be taken into account even in plastic facial surgery.

The observation that total human face impression is always dependent on present expressed facial emotion was first taken into consideration by Charles Darwin; Charles Darwin claimed there is a limited and universal set of facial emotions expressed by all higher mammals [8].

American psychologist Silvan Solomon Tomkins extended the idea by analyzing human facial emotions deeper in detail; he declared that specific facial expressions are uniquely linked to individual emotions and, not only but also, asserted that emotions are easily comprehensible across races, ethnic groups, and cultures [9].

In 1971, two psychologists Paul Ekman and Wallace Friesen established a classification of human facial impressions based on six (“clusters” of) emotions, (happiness, sadness, surprise, fear, anger, disgust) [10] and in 90’s they improved their classification of emotions by development of a well-known system called Facial Action Coding System (FACS). The system is based on movements of individual facial muscles which determine the resulting emotion in a form perceived by an observer [11].

In contradiction to the previous, also called *functional* approach how to classify human facial impressions into the appropriate emotions, there is another strategy, a *morphological* way – which is based on simple description of facial geometry [11].

Recognition techniques of human facial emotions come from a general human face image recognition techniques and can be divided into three phases [12]:

- (i) face detection and localization;
- (ii) extraction of appropriate face features;
- (iii) classification of a facial expression into a facial emotion.

The first phase, *face detection and localization*, could apply an *expert method* (e. g. left and right eye are both symmetric and of similar size, etc.) [13], a *feature invariant method* (e. g. eyes, nose, and mouth is detected by human perceiver regardless of an angle of view or intensity of current lighting), an *appearance-based method* (when face image is compared to face templates generated by a machine-learning algorithm) [14], [15].

The second phase, *extraction of appropriate face features*, can be done via *Gabor wavelets method* [16], an *image intensity analysis*, a *principal component analysis* (PCA), an *active appearance model* [17], or *graph models* [17], respectively, including also the well-known Marquardt mask.

Finally, the third phase, *classification of facial expression into emotion cluster*, is one of so-called *classification problem* and belongs to the families of machine-learning algorithms. It can be performed by rule-based classifiers [18], model-comparing classifiers [18] or machine-learning classifiers [18].

To conclude this up, aims of this study therefore are

- (i) to detect which facial geometric features and their surgical corrections are connected with increased facial attractiveness level in patients undergoing rhinoplasty;
- (ii) to work out and test a system of facial expressions based on FACS, so that it could be used for classification

of facial images into facial emotions – this could be a promising starting point for analysis of relations between facial expressions based on facial muscles geometry and movement, and facial emotions, respectively.

The second point (ii) seems to be crucial for planning of facial surgical procedures – whereas real structures such as facial muscles are already objects of surgical interventions, changes for the better in facial expressions should be in fact the desired results of the surgical procedures. However, the relations are not obvious, and machine-learning classification of facial emotions could be one of the first steps in the process of their clarification.

## II. RESEARCH MATERIAL AND METHODOLOGY

Patients who attended the Department of Plastic Surgery, First Faculty of Medicine, Charles University, Prague and Na Bulovce Hospital were asked to join the study and informed enough about all details of the study. There were precisely 30 patients in total who underwent the rhinoplasty surgery and were eligible to join the study. A portrait and profile picture of each of them was taken and stored in a secured database.

There is another sample of 12 patients (all of them are students at the Faculty of Biomedical Engineering, Czech Technical University in Prague) whose portrait and profile images were taken just at the moment they shew a facial expression according to the given incentive. An overview of the facial expressions is in Table I. The total number of their pictures is therefore equal to  $12 \times 14 = 168$ .

TABLE I  
RELATIONS BETWEEN THE FACIAL EMOTIONS AND THE THEIR QUALITY

facial emotion	quality
contact	positive
helpfulness	positive
evocation	positive
defence	negative
aggression	negative
reaction	neutral
decision	neutral
well-being	positive
fun	positive
rejection	negative
depression	negative
fear	negative
deliberation	positive
expectation	positive

*Data of Interests:* Besides the facial image data described one paragraph above, a seven-level Likert scale following the values of  $(-3, -2, -1, 0, +1, +2, +3)$  (the higher score, the more attractive is a face considered to be) was used to evaluate each photography of each patient before and after undergoing the rhinoplasty. There was a board of 14 independent evaluators doing the evaluation.

The facial emotions we used in the study are based on the Facial Action Coding System (FACS), but has been improved a bit. We defined 14 clusters of emotions in total – *contact*, *helpfulness*, *evocation*, *defence*, *aggression*, *reaction*, *decision*,

*well-being, fun, rejection, depression, fear, deliberation, and expectation*, respectively [19], [20].

Furthermore, we defined a *quality* of facial emotions such that each one of the emotions is either positive, negative or neutral, respectively, according to an average effect on a perceiver (and stated by an expert).

Relations between the facial emotions and the quality of the facial emotions, following the way how they were used in the study, are shown in Table I.

*Landmarking:* Landmarks can be defined as morphometrically essential points on a plane of a facial image. Overview of the landmarks of our interest is in the Fig. 1. Landmarks were plotted manually using proprietary program written in C#, by which the coordinates of all of them were collected. The landmarks were also obtained as well using an experimental application written in R language [21] which is able to bridge a well-known C++ library called `dlib` [22]; the `dlib` enables to use automatic facial landmarking. After the gathering of all landmarks' coordinates, for  $i$ -th landmark, where  $i \in \{1, 2, 3, \dots\}$ , with original coordinates  $[x_i, y_i]$ , new standardized coordinates  $[x'_i, y'_i]$  were computed in the terms of

$$x'_i = \frac{x_i - \min_{j \in \{1, 2, 3, \dots\}} \{x_j\}}{\max_{j \in \{1, 2, 3, \dots\}} \{x_j\} - \min_{j \in \{1, 2, 3, \dots\}} \{x_j\}}$$

$$y'_i = \frac{y_i - \min_{j \in \{1, 2, 3, \dots\}} \{y_j\}}{\max_{j \in \{1, 2, 3, \dots\}} \{y_j\} - \min_{j \in \{1, 2, 3, \dots\}} \{y_j\}},$$

assuming that all faces taken in the images are of equal size. The described transformation of coordinates (*standardization*) allows us to compare feasibly enough any two face portraits themselves (their transformed coordinates  $[x'_i, y'_i]$ ), and any two face profiles themselves (their transformed coordinates  $[x'_i, y'_i]$ ), respectively.

There are some of the derived metrics and angles calculated using the transformed coordinates of the landmarks in Table II (definitions of the landmarks are shown in Fig. 1).

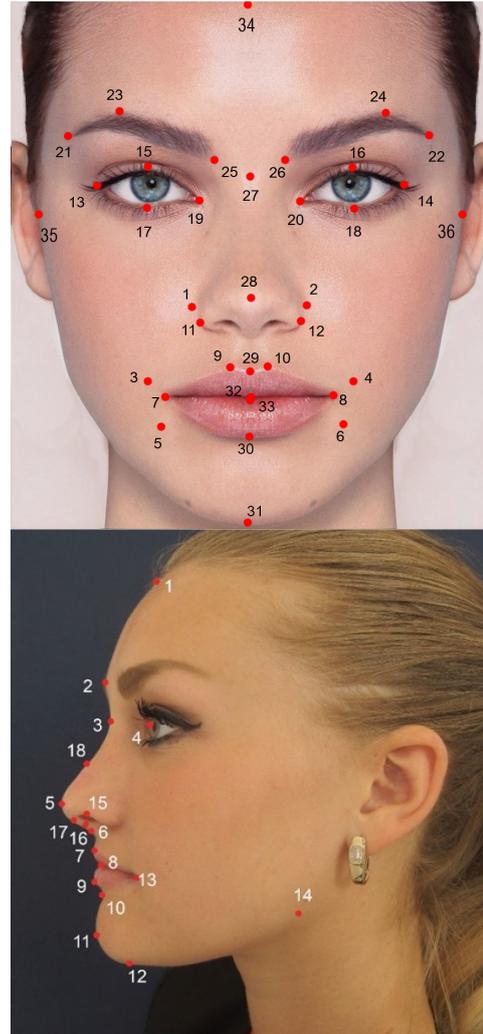


Fig. 1. Landmarks of a face portrait and a face profile

*Statistical Analysis:* Outputs with  $p$ -values below (or very “close” to) 0.05 were considered as statistically significant. All statistical analyses were performed using R language for statistical computing and graphics [21].

A multivariate regression analysis was carried out in order to identify which predictors, i. e. derived metrics or angles (see Table II for more details) statistically significantly affect an average difference of the attractiveness' Likert scores after and before the rhinoplasty undergoing [23].

Additionally, Bayesian naive classifiers [24], CART – classification and regression trees [25] and neural networks using backpropagation with sigmoidal activating function [26] were applied to classify an image of a human face (portrait) into one of the facial emotions, and as well into one of the levels of the quality of facial emotions (and even into some more parameters of emotions not discussed in this paper).

Performances of predictive accuracy of the previously mentioned three methods are reported as confusion matrices or as 95 % confidence intervals. Grant total sum of each of the confusion matrices is equal to  $12 \times 14 = 168$ , i. e. a number of

TABLE II  
SOME OF THE DERIVED METRICS AND ANGLES CALCULATED USING THE TRANSFORMED COORDINATES OF THE LANDMARKS

metrics/angles	definition
nasofrontal angle	angle between landmarks 2, 3, 18 (profile)
nasolabial angle	angle between landmarks 7, 6, 17 (profile)
nasal tip	horizontal Euclidean distance between landmarks 6, 5 (profile)
nostril prominence	Euclidean distance between landmarks 15, 16 (profile)
cornea-nasion distance	horizontal Euclidean distance between landmarks 3, 4 (profile)
outer eyebrow	Euclidean distance between landmarks 21, 22 (portrait)
inner eyebrow	Euclidean distance between landmarks 25, 26 (portrait)
lower lip	Euclidean distance between landmarks 30, 33 (portrait)
mouth height	Euclidean distance between landmarks 6, 8 (profile)
angular height	Euclidean distance between landmarks 7 (or 8) and 33 (portrait)

TABLE III  
SUMMARY OF THE MULTIVARIATE LINEAR REGRESSION

predictor	estimate	<i>t</i> -value	<i>p</i> -value
intercept <sub>after-before</sub>	3.832	1.696	0.043
nasofrontal angle <sub>after-before</sub>	0.353	0.174	0.050
nasolabial angle <sub>after-before</sub>	0.439	1.624	0.057
nasal tip <sub>after-before</sub>	-3.178	0.234	0.068
nostril prominence <sub>after-before</sub>	-0.145	0.128	0.266
cornea-nasion distance <sub>after-before</sub>	-0.014	0.035	0.694

individuals multiplied by a number of pictures per individual.

### III. RESULTS AND DISCUSSION

A summary of the multivariate linear regression is shown in Table III. As we can see from the Table III, the mean increase of facial attractiveness level after undergoing the rhinoplasty is about 3.8 Likert point,  $p = 0.043$ . Moreover, per each radian of nasofrontal angle enlargement, there is an expectation of mean increase about 0.353 Likert point in facial attractiveness after undergoing the rhinoplasty (when a patient went through this kind of correction),  $p = 0.050$ . Similarly, per each radian of nasolabial angle enlargement, there is an expectation of mean increase about 0.439 Likert point in facial attractiveness after undergoing the rhinoplasty (again, this can be true if and only if this kind of correction is even applied to a patient),  $p = 0.057$ .

As we expected, the larger both nasofrontal and nasolabial angles corrections are, the higher score of attractiveness level such a face obtains. Furthermore, the two mentioned angles are the main corrections which could be done within a routine rhinoplasty procedure. Of course, these results are limited. For instance, if both angles, nasofrontal and nasolabial one would be considered as straight angles, a nose would “disappear” under these conditions instead of expected facial attractiveness level increasing as stated above.

There are confusion matrices of the prediction of the emotional quality based both on Bayesian naive classifier (Table IV) and neural network (Table V). Confusion matrices of the prediction of the facial emotions are not reported due to the fact they oversize the page format.

Point estimate and 95 % confidence interval of mean prediction accuracy of the facial emotions based on Bayesian naive classifier is 0.325 (0.321, 0.329). Point estimate and 95 % confidence interval of mean prediction accuracy of

TABLE IV  
CONFUSION MATRIX OF A PREDICTION OF THE EMOTIONAL QUALITY BASED ON BAYESIAN NAIVE CLASSIFIER

		predicted class		
		negative	neutral	positive
true class	negative	34	16	16
	neutral	11	39	8
	positive	4	10	30

TABLE V  
CONFUSION MATRIX OF A PREDICTION OF THE EMOTIONAL QUALITY BASED ON A NEURAL NETWORK

		predicted class		
		negative	neutral	positive
true class	negative	36	6	6
	neutral	12	54	16
	positive	2	4	32

the emotional quality based on Bayesian naive classifier is 0.413 (0.409, 0.417). Since  $0.325 > \frac{1}{|\text{clusters of emotions}|} = \frac{1}{14}$  and  $0.413 > \frac{1}{|\text{emotional quality}|} = \frac{1}{3}$ , both classifiers predict more precise than random process. Since the target variables (*facial emotions* and *quality of facial emotions*, respectively) contain multiple classes, the classification task here is so-called “multiclass” and even only moderate prediction accuracy is acceptable under this conditions [27], [28].

Point estimate and 95 % confidence interval of mean prediction accuracy of the facial emotions based on decision trees is 0.488 (0.484, 0.492). Point estimate and 95 % confidence interval of mean prediction accuracy of the emotional quality based on decision trees is 0.525 (0.521, 0.529). Similarly, in both cases, the classifier predicts more precise than a random process.

Finally, point estimate and 95 % confidence interval of mean prediction accuracy of the facial emotions based on neural networks is 0.507 (0.503, 0.511). Point estimate and 95 % confidence interval of mean prediction accuracy of the emotional quality based on neural network is 0.726 (0.722, 0.730). Again, in both cases, the classifier predicts far more precise than a random process (and even substantially better than the previous two classifiers, though).

There are examples of decision trees learned in order to predict one of the facial emotions or one of the emotional quality using facial geometry of the photographed facial ex-

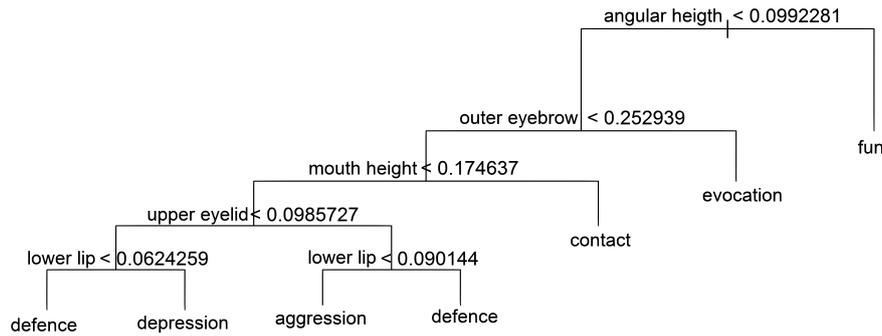


Fig. 2. A decision tree for prediction of the facial emotions (statements in nodes are true for left child nodes)

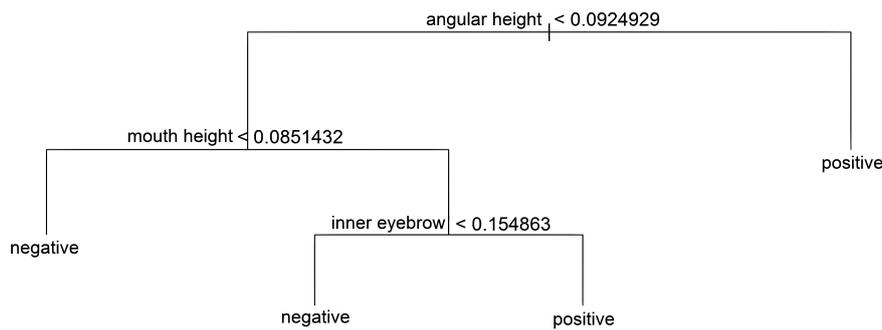


Fig. 3. A decision tree for prediction of the quality of facial emotions (statements in nodes are true for left child nodes)

pression in Fig. 2 and Fig. 3. The closer to the root node the derived geometrical metric or angle in the plot is, the more important seems to be in order to explain a “direction” of the classification into the final class of interest. As we can see, the facial expressions are dominated by geometry of the mouth, then by geometry of the eyes, respectively.

Furthermore, once we would go deeper into results of Fig. 2, we could realize that if the angular height — that is a vertical distance between mouth angles and a horizontal line between the lips — is large enough (more precisely, if the angular height is larger than or eventually equal to 0.0992), and it means that such a face in the image is smiling, then an emotion of that image is classified as a fun, as we can expect. Similar derivations (and still feasible) could be done following the “rules” placed in the next nodes of the trees plotted in Fig. 2 and Fig. 3.

#### IV. CONCLUSION

The performed machine-learning analyses pointed out which geometric facial features, based on significant data evidence, affect facial attractiveness the most — either as predictors increasing facial attractiveness level after undergoing rhinoplasty or as geometric features influencing the classification of facial images into facial emotions —, and therefore should preferentially be treated within rhinoplasty procedures.

Moreover, the learned classification methods confirmed that they are, despite the suggested improvement of FACS scale in terms of increasing the number of facial emotions, able

to classify facial images into the defined facial emotions accurately enough.

#### V. CONFLICT OF INTEREST

The authors declare that they have no conflict of interest regarding the publication of this article.

#### REFERENCES

- [1] Leslie G. Farkas, Tania A. Hreczko, John C. Kolar, et al. “Vertical and Horizontal Proportions of the Face in Young Adult North American Caucasians”. In: *Plastic and Reconstructive Surgery* 75.3 (Mar. 1985), pp. 328–337. DOI: 10.1097/00006534-198503000-00005. URL: <https://doi.org/10.1097/00006534-198503000-00005>.
- [2] Kendra Schmid, David Marx, and Ashok Samal. “Computation of a face attractiveness index based on neoclassical canons, symmetry, and golden ratios”. In: *Pattern Recognition* 41.8 (Aug. 2008), pp. 2710–2717. DOI: 10.1016/j.patcog.2007.11.022. URL: <https://doi.org/10.1016/j.patcog.2007.11.022>.
- [3] Mounir Bashour. *Is an objective measuring system for facial attractiveness possible*. Boca Raton, Fla: Dissertation.com, 2007. ISBN: 978-158-1123-654.
- [4] Randy Thornhill and Steven W. Gangestad. “Human facial beauty”. In: *Human Nature* 4.3 (Sept. 1993), pp. 237–269. DOI: 10.1007/bf02692201. URL: <https://doi.org/10.1007/bf02692201>.

- [5] A. C. Little, B. C. Jones, and L. M. DeBruine. “Facial attractiveness: evolutionary based research”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 366.1571 (May 2011), pp. 1638–1659. DOI: 10.1098/rstb.2010.0404. URL: <https://doi.org/10.1098/rstb.2010.0404>.
- [6] D. I. Perrett, K. J. Lee, I. Penton-Voak, et al. “Effects of sexual dimorphism on facial attractiveness”. In: *Nature* 394.6696 (Aug. 1998), pp. 884–887. DOI: 10.1038/29772. URL: <https://doi.org/10.1038/29772>.
- [7] Farhad Naini. *Facial aesthetics : concepts & clinical diagnosis*. Chichester, West Sussex, UK Ames, Iowa: Wiley-Blackwell, 2011. ISBN: 978-1-405-18192-1.
- [8] Charles Darwin. *The expression of the emotions in man and animals*. Oxford New York: Oxford University Press, 1998. ISBN: 9780195158069.
- [9] Silvan Tomkins. *Affect imagery consciousness : the complete edition*. New York: Springer Pub, 2008. ISBN: 978-0826144041.
- [10] Paul Ekman and Wallace V. Friesen. “Constants across cultures in the face and emotion.” In: *Journal of Personality and Social Psychology* 17.2 (1971), pp. 124–129. DOI: 10.1037/h0030377. URL: <https://doi.org/10.1037/h0030377>.
- [11] Paul Ekman. *Unmasking the face : a guide to recognizing emotions from facial clues*. Cambridge, MA: Malor Books, 2003. ISBN: 1883536367.
- [12] B. Fasel and Juergen Luetttin. “Automatic facial expression analysis: a survey”. In: *Pattern Recognition* 36.1 (Jan. 2003), pp. 259–275. DOI: 10.1016/s0031-3203(02)00052-3. URL: [https://doi.org/10.1016/s0031-3203\(02\)00052-3](https://doi.org/10.1016/s0031-3203(02)00052-3).
- [13] Ming-Hsuan Yang, D.J. Kriegman, and N. Ahuja. “Detecting faces in images: a survey”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24.1 (2002), pp. 34–58. DOI: 10.1109/34.982883. URL: <https://doi.org/10.1109/34.982883>.
- [14] A Lanitis, CJ Taylor, and TF Cootes. “Automatic face identification system using flexible appearance models”. In: *Image and Vision Computing* 13.5 (June 1995), pp. 393–401. DOI: 10.1016/0262-8856(95)99726-h. URL: [https://doi.org/10.1016/0262-8856\(95\)99726-h](https://doi.org/10.1016/0262-8856(95)99726-h).
- [15] Henry A. Rowley, Shumeet Baluja, and Takeo Kanade. “Neural Network-Based Face Detection”. In: *IEEE Trans. Pattern Anal. Mach. Intell.* 20.1 (Jan. 1998), pp. 23–38. ISSN: 0162-8828. DOI: 10.1109/34.655647. URL: <http://dx.doi.org/10.1109/34.655647>.
- [16] Xiaoming Zhao and Shiqing Zhang. “A Review on Facial Expression Recognition: Feature Extraction and Classification”. In: *IETE Technical Review* 33.5 (Jan. 2016), pp. 505–517. DOI: 10.1080/02564602.2015.1117403. URL: <https://doi.org/10.1080/02564602.2015.1117403>.
- [17] T.F. Cootes, C.J. Taylor, D.H. Cooper, et al. “Active Shape Models-Their Training and Application”. In: *Computer Vision and Image Understanding* 61.1 (Jan. 1995), pp. 38–59. DOI: 10.1006/cviu.1995.1004. URL: <https://doi.org/10.1006/cviu.1995.1004>.
- [18] Ethem Alpaydin. *Introduction to machine learning*. Cambridge, Mass: MIT Press, 2010. ISBN: 9780262012430.
- [19] Pavel Kasal, Patrik Fiala, Lubomír Štěpánek, et al. “Application of Image Analysis for Clinical Evaluation of Facial Structures”. In: *Medsoft 2015* (2015), pp. 64–70. URL: [http://www.creativeconnections.cz/medsoft/2015/Medsoft\\_2015\\_kasal.pdf](http://www.creativeconnections.cz/medsoft/2015/Medsoft_2015_kasal.pdf).
- [20] Lubomir Stepanek, Pavel Kasal, and Jan Mestak. “Evaluation of facial attractiveness for purposes of plastic surgery using machine-learning methods and image analysis”. In: *20th IEEE International Conference on e-Health Networking, Applications and Services, Healthcom 2018, Ostrava, Czech Republic, September 17-20, 2018*. 2018, pp. 1–6. DOI: 10.1109/HealthCom.2018.8531195. URL: <https://doi.org/10.1109/HealthCom.2018.8531195>.
- [21] R Core Team. *R: A Language and Environment for Statistical Computing*. ISBN 3-900051-07-0. R Foundation for Statistical Computing. Vienna, Austria, 2013. URL: <http://www.R-project.org/>.
- [22] Davis E. King. “Dlib-ml: A Machine Learning Toolkit”. In: *J. Mach. Learn. Res.* 10 (Dec. 2009), pp. 1755–1758. ISSN: 1532-4435. URL: <http://dl.acm.org/citation.cfm?id=1577069.1755843>.
- [23] John Chambers. *Statistical models in S*. Boca Raton, Fla: Chapman & Hall/CRC, 1992. ISBN: 041283040X.
- [24] Nir Friedman, Dan Geiger, and Moises Goldszmidt. “Bayesian Network Classifiers”. In: *Mach. Learn.* 29.2-3 (Nov. 1997), pp. 131–163. ISSN: 0885-6125. DOI: 10.1023/A:1007465528199. URL: <https://doi.org/10.1023/A:1007465528199>.
- [25] Leo Breiman. *Classification and regression trees*. New York: Chapman & Hall, 1993. ISBN: 0412048418.
- [26] Warren S. McCulloch and Walter Pitts. “A logical calculus of the ideas immanent in nervous activity”. In: *The Bulletin of Mathematical Biophysics* 5.4 (Dec. 1943), pp. 115–133. DOI: 10.1007/bf02478259. URL: <https://doi.org/10.1007/bf02478259>.
- [27] Johannes Stallkamp, Marc Schlipsing, Jan Salmen, et al. “The German Traffic Sign Recognition Benchmark: A multi-class classification competition”. In: *The 2011 International Joint Conference on Neural Networks*. IEEE, July 2011. DOI: 10.1109/ijcnn.2011.6033395. URL: <https://doi.org/10.1109/ijcnn.2011.6033395>.
- [28] Sridhar Ramaswamy, Pablo Tamayo, Ryan Rifkin, et al. “Multiclass cancer diagnosis using tumor gene expression signatures”. In: *Proceedings of the National Academy of Sciences* 98.26 (2001), pp. 15149–15154. ISSN: 0027-8424. DOI: 10.1073/pnas.211566398. eprint: <https://www.pnas.org/content/98/26/15149.full.pdf>. URL: <https://www.pnas.org/content/98/26/15149>.

# 12<sup>th</sup> International Workshop on Computational Optimization

**M**ANY real world problems arising in engineering, economics, medicine and other domains can be formulated as optimization tasks. These problems are frequently characterized by non-convex, non-differentiable, discontinuous, noisy or dynamic objective functions and constraints which ask for adequate computational methods.

The aim of this workshop is to stimulate the communication between researchers working on different fields of optimization and practitioners who need reliable and efficient computational optimization methods.

## TOPICS

The list of topics includes, but is not limited to:

- combinatorial and continuous global optimization
- unconstrained and constrained optimization
- multiobjective and robust optimization
- optimization in dynamic and/or noisy environments
- optimization on graphs
- large-scale optimization, in parallel and distributed computational environments
- meta-heuristics for optimization, nature-inspired approaches and any other derivative-free methods
- exact/heuristic hybrid methods, involving natural computing techniques and other global and local optimization methods
- numerical and heuristic methods for modeling

The applications of interest are included in the list below, but are not limited to:

- classical operational research problems (knapsack, traveling salesman, etc)
- computational biology and distance geometry
- data mining and knowledge discovery
- human motion simulations; crowd simulations
- industrial applications
- optimization in statistics, econometrics, finance, physics, chemistry, biology, medicine, and engineering
- environment modeling and optimization

## BEST PAPER AWARD

The best WCO'19 paper will be awarded during the social dinner of FedCSIS 2019.

The best paper will be selected by WCO'19 co-Chairs by taking into consideration the scores suggested by the reviewers, as well as the quality of the given oral presentation.

## EVENT CHAIRS

- **Fidanova, Stefka**, Bulgarian Academy of Sciences, Bulgaria
- **Mucherino, Antonio**, INRIA, France
- **Zaharie, Daniela**, West University of Timisoara, Romania

## PROGRAM COMMITTEE

- **Abud, Germano**, Universidade Federal de Uberlândia, Brazil
- **Andrei, Stefan**
- **Bonates, Tibérius**, Universidade Federal do Ceará, Brazil
- **Breaban, Mihaela**
- **Gruber, Aritanan**
- **hadj salem, khadija**, University of Tours - LIFAT Laboratory, France
- **Hosobe, Hiroshi**, Hosei University, Japan
- **Lavor, Carlile**, IMECC-UNICAMP, Brazil
- **Marin, Mircea**
- **Micota, Flavia**, West University of Timisora, Romania
- **Muscalagiu, Ionel**, Politehnica University Timisoara, Romania
- **Pintea, Camelia**, Tehnical University Cluj-Napoca, Romania
- **Stefanov, Stefan**, South-West University Neofit Rilski, Bulgaria
- **Stoian, Catalin**, University of Craiova, Romania
- **Wang, Yifei**
- **Zilinskas, Antanas**, Vilnius University, Lithuania



# A Minimum Set-Cover Problem with several constraints

Jens Dörpinghaus\*, Carsten Düing†, Vera Weil‡

Fraunhofer Institute for Algorithms and Scientific Computing,  
Schloss Birlinghoven, Sankt Augustin, Germany

Email: \*jens.doerpinghaus@scai.fraunhofer.de, †carsten.cdueing@scai.fraunhofer.de

‡Department for Computer Science,  
University of Cologne, Germany

Email: weil@informatik.uni-koeln.de

**Abstract**—A lot of problems in natural language processing can be interpreted using structures from discrete mathematics. In this paper we will discuss the search query and topic finding problem using a generic context-based approach. This problem can be described as a Minimum Set Cover Problem with several constraints. The goal is to find a minimum covering of documents with the given context for a fixed weight function. The aim of this problem reformulation is a deeper understanding of both the hierarchical problem using union and cut as well as the non-hierarchical problem using the union. We thus choose a modeling using bipartite graphs and suggest a novel reformulation using an integer linear program as well as novel graph-theoretic approaches.

## I. INTRODUCTION

In scientific research, expert systems provide users with several methods for knowledge discovery. They are widely used to find relevant or novel information. For example, medical and biological researchers try to find molecular pathways, mechanisms within living organisms or special occurrences of drugs or diseases. In [1], we discussed a novel approach for describing NLP problems using theoretical computer science. Using this approach, it is possible to obtain the algorithmic core of a NLP problem. Here, we will discuss two  $\mathcal{NP}$ -complete problems: Search Query Finding (SQF) and Topic Finding (TF).

Using expert system as an input, researches usually consider an initial idea and some content like papers or other documents. The most common approach is inquiring a search engine to find closely related information. Thus two question are most frequently asked: "How can I find these documents?" to adjust the search query for knowledge discovery or "What are these documents all about?" to find the topic. Both questions are heavily related to the context of documents. Metadata like authors, keywords and text are used to retrieve results of a query using a search engine.

Semantic searches are usually based on textual data and some meta-data like authors, journals, keywords. In addition, time and complexity play an important role, since often relevant information is not findable or new information is already available. For example, databases such as PubMed [2]

contain around 27 million abstracts and PMC<sup>1</sup> includes around 2 million biomedical-related full-text articles.

Both problems are equivalent (see [1]) and can be described as a Minimum Set Cover Problem with several constraints. Query languages and natural languages are not only highly connected but merge more and more (see [3] or [4]). The goal is to find a minimum covering of documents with the given context for a fixed weight function. The aim of this problem reformulation is a deeper understanding of both the hierarchical as well as the non-hierarchical problem. We thus choose a modeling using bipartite graphs and suggest a novel reformulation using an integer linear program as well as graph-theoretic approaches.

There is a considerable amount of literature on both problems. Many studies have been published on probabilistic or machine-learning-approaches, see [5], [6] or [7]. In addition, in recent years there has been growing interest in providing users with suggestions for more specific or related search queries, see [8].

This paper is divided into six sections. The first section gives a brief overview of the problem formulation and provides the definition of MDC and WMDC. The second section analyses the hierarchical problem formulation and proposes novel heuristics. In the third section, we present a short analysis of the non-hierarchical problem and propose an integer linear program approach and some modified graph heuristics to solve this problem. We present some experimental results on artificial and real-world scenarios in section four. Our conclusions are drawn in the final section.

## II. PROBLEM FORMULATION AND DEFINITION

We follow the notation introduced in [1]. Let  $\mathbb{D}$  be a set of documents and let  $\mathbb{X}$  be a set of context data. Context data is information associated with documents, such as keywords, authors, publication venue, etc. Both  $\mathbb{D}$  and  $\mathbb{X}$  form the vertex set of a graph  $G$ . If and only if a description of a document  $d \in \mathbb{D}$  is associated with context data  $x \in \mathbb{X}$ , we add the edge  $\{d, x\}$  to  $E$ . The graph  $G = (\mathbb{D} \cup \mathbb{X}, E)$  is bipartite and called *document description graph*.

<sup>1</sup><https://www.ncbi.nlm.nih.gov/pmc/>

Given a subset  $R \subset \mathbb{D}$ , the search-query-finding (SQF) or topic-finding (TF) problem tries to find a good description of  $R$  with terms in  $\mathbb{X}$ . In general, we lack a proper definition of what *good* means.

For example, given a search engine  $q : \mathbb{X} \rightarrow \mathbb{D}$  and a description function  $f : \mathbb{D} \rightarrow \mathbb{X}$ , we want a solution  $Z \subset \mathbb{X}$  such that  $q(Z) = R$  and  $Z = f(R)$ . If we want to obtain a human-readable topic for  $R$ , we need a solution  $Z$  of minimum cardinality which precisely describes all documents in  $R$ , hence distinguishing  $R$  from  $\mathbb{D} \setminus R$  without duplication and redundancies. See Figure 1 for an illustration of the relation between the sets  $X, R$  and the mappings  $f, q$ .

To sum up, we need to find a minimum covering of  $R$  with elements in  $\mathbb{X}$  so that whenever we are forced to cover further documents, that is, documents in  $\mathbb{D} \setminus R$ , the number of these further documents is minimal. Depending on the considered problem and the usecase, we have to make a trade-off between the size of the subset in  $\mathbb{X}$  and the number of covered documents in  $\mathbb{D} \setminus R$ . However, these problems are all related to the problem of finding dominating sets in bipartite graphs, see [9]. The latter is  $\mathcal{NP}$ -complete, even for bipartite graphs, see [10].

For  $x_i \in \mathbb{X}$ , we call  $D_i = N(x_i) \subseteq \mathbb{D}$  the cover set of  $x_i$  in  $\mathbb{D}$ . Roughly speaking, just imagine a keyword  $x_i$  and all associated documents  $D_i$ . With this, we reformulate the problem as follows:

**Definition II.1.** (*Document Cover Problem, DC*) Let  $\mathbb{D}$  be a set of documents, let  $\mathbb{X}$  be a set of context data and let  $G = (\mathbb{D} \cup \mathbb{X}, E)$  be the document description graph.

Given a set of documents  $R \subset \mathbb{D}$ , a solution of the DC is a set  $C \subseteq D$  that covers at least  $R$ .

**Definition II.2.** (*Minimum Document Cover Problem, MDC*) Let  $C$  be a solution of the DC and let  $\alpha_2 = |C|$ . Let further  $\alpha_1 = r$  be the number of documents in  $C \setminus R$ .

A solution of MDC is a solution of DC so that  $\alpha = \alpha_1 + \alpha_2$  is minimal.

We can define two objectives for minimization:  $\alpha_1$  and  $\alpha_2$ .

**Definition II.3.** ( $\alpha_2$ -Minimum Document Cover Problem,  $\alpha_2$ -MDC) Given a set of documents  $R \subset \mathbb{D}$ , a solution of the  $\alpha_2$ -MDC is a solution of DC so that  $\alpha = \alpha_2$  is minimal.

**Definition II.4.** ( $\alpha_1$ -Minimum Document Cover Problem,  $\alpha_1$ -MDC) Given a set of documents  $R \subset \mathbb{D}$ , a solution of the  $\alpha_1$ -MDC is a solution  $C$  of DC so that  $\alpha = \alpha_1$  is minimal.

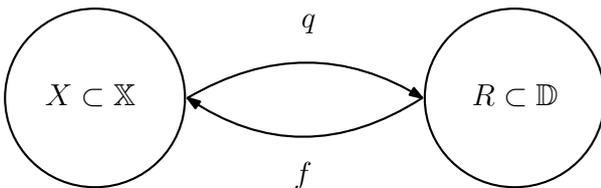


Fig. 1: Relation between the sets  $X \subset \mathbb{X}$  as description set of documents in  $R \subset \mathbb{D}$ .

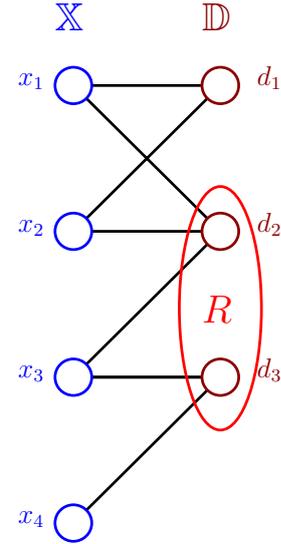


Fig. 2: A graph  $G = (\mathbb{D} \cup \mathbb{X}, E)$  illustrating example II.8

We further introduce a weighted version of this problem:

**Definition II.5.** (*Weighted Minimum Document Cover Problem, WMDC*) Let  $\mathbb{D}$  be a set of documents, let  $\mathbb{X}$  be a set of context data and let  $G = (\mathbb{D} \cup \mathbb{X}, E)$  the document description graph. Let  $w : \mathbb{X} \rightarrow \mathbb{R}$  be a weight function which associates a weight for every element in  $\mathbb{X}$ . Moreover, we set  $D = \{D_1, \dots, D_n\}$ . Let  $\alpha_1 = r$  be the number of documents in  $R \subset \mathbb{D}$  and  $\alpha_2 = |C|$ .

A solution of the WMDC is a set  $C \subseteq D$  which covers  $R$ , such that  $\alpha = \alpha_1 + \alpha_2 + w(C)$  is minimal, where  $w(C) = \sum_{c \in C} w(c)$ .

Again we can find formulations for  $\alpha_1$ -WMDC and  $\alpha_2$ -WMDC. Both problems are  $\mathcal{NP}$ -hard, see [11].

In general, we will focus on the  $\alpha_2$  optimization. Thus, in this paper, we denote this version with the MDC and WMDC.

We have to distinguish between hierarchical and non-hierarchical approaches. Both MDC and WMDC search for a cover set  $c_1, \dots, c_n$  which leads to a solution  $c_1 \cup \dots \cup c_n$ . This is a non-hierarchical approach. Using a search engine this would lead to a solution  $c_1$  or  $\dots$  or  $c_n$ . Utilizing the cut of sets we will need a hierarchical solution  $(c_1 \cup \dots \cup c_n) \cap (c_{n+1} \cup \dots \cup c_m) \cap \dots$ . Using a search engine would lead to a solution  $(c_1$  or  $\dots$  or  $c_n)$  and  $(c_{n+1}$  or  $\dots$  or  $c_m)$  and ....

**Definition II.6.** (*Hierarchical Minimum Document-Cover Problem, HMDC*) Let  $\mathbb{D}$  be a set of documents, let  $\mathbb{X}$  be a set of context data and let  $G = (\mathbb{D} \cup \mathbb{X}, E)$  be the document description graph. Moreover, we set  $D = \{D_1, \dots, D_n\}$ .

A solution of the HMDC problem for  $R \subset \mathbb{D}$  is a minimum cover  $C \subseteq D$  with  $C = C_1 \cap \dots \cap C_n$  and  $C_i = C_1^i \cup \dots \cup C_m^i$  of  $R$  so that  $C \setminus R$  is minimal. We use  $N(x_i)$  as usual for the open neighborhood  $N(x_i) \setminus x_i$ .

**Definition II.7.** (*Hierarchical Weighted Minimum Document-*

*Cover Problem, HWMDP*) Given a set of documents  $\mathbb{D}$ , a set of context data  $\mathbb{X}$  and the document description graph  $G = (\mathbb{D} \cup \mathbb{X}, E)$ . We set  $D = \{D_1, \dots, D_n\}$ . Given a weight function  $w : \mathbb{X} \rightarrow \mathbb{R}$  that defines a weight for every element in  $\mathbb{X}$ .

A solution of the weighted HWMDP problem for  $R \subset \mathbb{D}$  is a minimum cover  $C \subseteq D$  with  $C = C_1 \cap \dots \cap C_n$  and  $C_i = C_1^i \cup \dots \cup C_m^i$  of  $R$ , i.e.  $\sum_{c \in C} w(c)$  is minimal, so that  $C \setminus R$  is minimal.

We will discuss two examples for the non-hierarchical problem:

**Example II.8.** Given an instance of the MDC with  $\mathbb{D} = \{d_1, d_2, d_3\}$ ,  $R = \{d_2, d_3\}$ ,  $\mathbb{X} = \{x_1, \dots, x_4\}$  and  $D_1 = D_2 = \{d_1, d_2\}$ ,  $D_3 = \{d_2, d_3\}$ ,  $D_4 = \{d_3\}$ . See figure 2 for an illustration.

A minimum set cover cannot include  $x_1$  or  $x_2$ , but a solution is  $C = D_3$ .

**Example II.9.** Consider the instance given in example II.8 with additional weights  $w(x_1) = w(x_2) = w(x_3) = 1$  and  $w(x_4) = 0$ . A minimum solution of the weighted MDC can be found with  $Z = \{x_3, x_4\}$ .

Let  $w(x_1) = w(x_3) = 1$  and  $w(x_4) = w(x_2) = 0$ . A minimum solution of weighted MDC can be either found with  $Z = \{x_2, x_4\}$ , here  $w(Z) = 0$  but  $|C \setminus R| = 1$ . If we chose  $Z = \{x_3, x_4\}$   $w(Z) = 1$  but  $|C \setminus R| = 0$ .

We will first of all focus on hierarchical approaches, discussing approaches using dynamic programming and bipartite graph heuristics or spanning trees. After that we will discuss the non-hierarchical problem and solutions using an integer linear program approach as well as some heuristics utilizing the graph structure. We will evaluate the results on some random instances and finish with a conclusion.

### III. HIERARCHICAL APPROACHES

#### A. Problem Description

For some questions it is interesting to find a cover of  $R \subset \mathbb{D}$  with increasing (decreasing) or selectable exactness and the number of named entities  $Z \subset X = f(R)$ . If we have a set of documents and want to obtain more others closely related documents, we may be interested in a modification of the similarity measure for documents or search queries. We build covers  $C_i = q(Z_i)$  of  $R$  and optimize the solution by concatenating them with a logical AND.

#### B. Using unique keyword descriptions on bipartite graphs

From the graph in figure 2 we can see that the graph  $G = (\mathbb{D} \cup \mathbb{X}, E)$  is bipartite. The neighborhood  $N(d) \subset \mathbb{X}$  of every document  $d \in \mathbb{D}$  is not necessarily unique description of this document. Thus we can find a trivial solution of the MDCP on  $R \subset \mathbb{D}$  by

$$\bigvee_{d \in R} (\bigwedge_{x \in N(d)} x)$$

We can eliminate elements with the largest error from this list. This process can be limited by iterations as well as a

precision. For example we may limit the precision to 0.9 which will eliminate at maximum 10% of all keywords, whereas a precision of 0.5 will eliminate at maximum 50%.

---

#### Algorithm 1 KEYWORD-COVER

---

**Require:** Documents  $\{d_1, \dots, d_n\} \subset \mathbb{D}$  and descriptive elements  $f(d_i) = \{x_1, \dots, x_m\} \subset \mathbb{X}$ , a weight function  $w : \mathbb{X} \rightarrow \mathbb{R}$  maxiter as maximum of iterations, prec as precision

**Ensure:** A cover  $Z = (x_i \wedge x_j \wedge \dots) \vee (x_k \wedge x_l \wedge \dots) \vee \dots$  of  $R$  with elements in  $\mathbb{X}$ .

```

1:  $f' = f$ 
2: for every  $d \in R$  do
   while iteration < maxiter AND  $f'(d) > (\text{prec} \cdot f(d))$ 
   do
3:   remove  $x \in f'(d)$  with maximum weight
   end while
6: end for
   return  $Z = \bigvee_{d \in R} (\bigwedge_{x \in f'(d)} x)$ 

```

---

If we set  $w : \mathbb{X} \rightarrow \mathbb{R}$  as the error function  $\text{err}(x) = |q(x) \setminus R|$  we will find a solution for MDCP, otherwise this will return a solution of WMDCP. The function  $\text{err}$  is a less time-consuming approach but highly depended on the distribution of  $\mathbb{X}$ .

#### C. Dynamic programming and bipartite graph heuristic

Here, we describe a heuristic and dynamic method by creating dominating subgraphs of a bipartite graph. Building the bipartite graph  $G_b = (V = R \cup X, E)$ , a subgraph of the document description graph  $G = (\mathbb{D} \cup \mathbb{X}, E)$ , we create a set with documents  $R_a = \{d_1, \dots, d_n\} \subseteq \mathbb{D}$  and all their context data (like keywords, named entities, etc.) in a sorted list  $X_a = \{x_1, \dots, x_m\} \subseteq \mathbb{X}$  for the two sets of nodes. The edges  $(d_i, x_j)$  in  $G_b$  are given for all pairs  $d_i, x_j$  iff  $x_j \in f(d_i)$ . The elements in  $X_a$  should be sorted ascending or descending by their degree. For our example we choose a descending order, which results in an increasing precise cover.

In addition we need to build a second set  $R_b$  as temporary storage for the documents and a sorted list of lists  $Z = \{Z_1, Z_2, \dots, Z_k\}$ , with the covers  $Z_i$  of  $R_a$  for the output. The algorithm in pseudocode can be found in alg. 2. In every execution of the while loop in line 7 a new sublist  $Z_i \subset Z$  is created (see line 13). All of them are complete covers of all documents in  $R_a$ , where  $Z_0$  may contain just one element  $x_i$  with  $N(x_i) = R_a$  and the last  $Z_m$  may contain just all identities, that means  $x_i$  with a single neighbor  $N(x_i) = d_i$ . There are many options to modify the algorithm for special use cases. Choosing the ascending order for  $X_a$  and the minimum in line 9, which is same as in the other case just means the first  $x_j \in X_a$ , will mostly give different results.

If after the last run of the loop  $X_a$  is empty, but there are still documents in  $R_a$ , we receive an incomplete cover  $Z_k$ . To avoid that we add the ID's for the last documents in  $R_a$  (in descending order) to  $Z_k$ , or create and add an all covering  $x_\infty$  (for descending order).



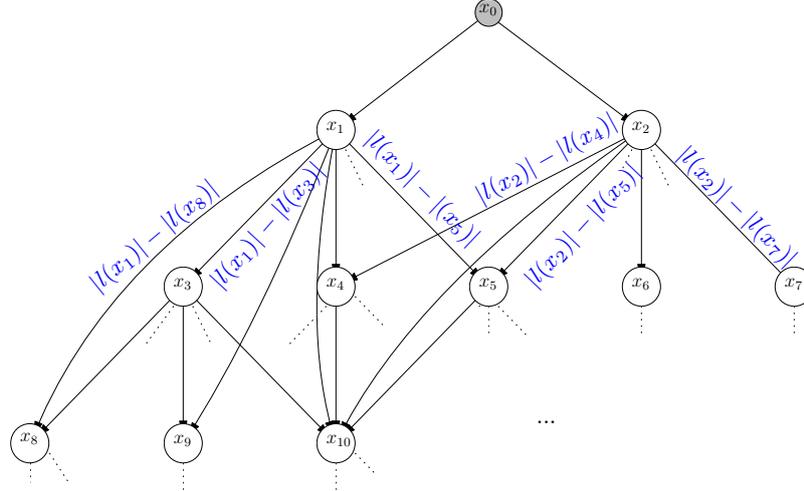


Fig. 4: Illustration of set representative in the graph  $G' = (V, E)$  and weight  $w(x_i, x_j)$  after adding the meta node  $x_0$ , with  $l(x_i) := |N_G(x_i)|$ . Not all edges and nodes have been added.

### B. An Integer Linear Program Approach

Numerous ILP-formulations for the set-cover problem can be found in literature, for example [13] or [14]. To meet definition II.6 of MDC we need to adjust the formulation.

Given a set of documents  $\mathbb{D}$ , a subset  $R \subset \mathbb{D}$ , a set of context data  $f(R) = X \subset \mathbb{X}$  and the document description graph  $G = (\mathbb{D} \cup \mathbb{X}, E)$ . We can define  $\forall x_i \in \mathbb{X} D_i = N(x_i)$  as the cover set of  $x_i$  in  $\mathbb{D}$ . We set  $D = \{D_1, \dots, D_n\}$  and  $e(D_i) = D_i \setminus R$  as the error of the description term  $x_i$ .

A solution of the MDC problem for  $R \subset \mathbb{D}$  is a minimum cover  $C \subseteq D$  of  $R$  so that  $C \setminus R$  is minimal.

$$\begin{aligned} \min \quad & \sum_{i=1}^n x_i + \sum_{i=1}^n x_i e(X_i) \\ \text{subject to} \quad & \sum_{i:v \in X_i} x_i \geq 1, \forall v \in R \\ & x_i \geq 1 \quad \forall i = 1, \dots, n \\ & x_i \in \mathbb{Z} \quad \forall i = 1, \dots, n \end{aligned} \quad (1)$$

Here the vector  $x$  gives a set  $Z \subset X$  which gives a minimum cover  $q(Z) = C \subseteq D$  of  $R$  so that  $C \setminus R$  is minimal.

The weighted MDC problem was introduced in definition II.7. Given a weight function  $w : \mathbb{X} \rightarrow \mathbb{R}$  that defines a weight for every element in  $\mathbb{X}$  the ILP (1) changes as follows:

$$\begin{aligned} \min \quad & \sum_{i=1}^n w(x_i) + \sum_{i=1}^n x_i e(X_i) \\ \text{subject to} \quad & \sum_{i:v \in X_i} x_i \geq 1 \quad \forall v \in R \\ & x_i \geq 1 \quad \forall i = 1, \dots, n \\ & x_i \in \mathbb{Z} \quad \forall i = 1, \dots, n \end{aligned} \quad (2)$$

A solution of the MDC problem for  $R \subset \mathbb{D}$  is a minimum cover  $C \subseteq D$  of  $R$ , i.e.  $\sum_{c \in C} w(c)$  is minimal, so that  $C \setminus D$  is minimal.

### C. Dynamic programming and bipartite graph heuristic

We can use algorithm 2 to construct a non-hierarchical solution. This algorithm has already been used to compute  $k$  covers of  $R_a$ , which can be used to find a cover with minimal

---

### Algorithm 4 BIPARTITE COVER-DESCRIPTION

---

**Require:** Documents  $\{d_1, \dots, d_n\} \subset \mathbb{D}$  and descriptive elements  $f(d_i) = \{x_1, \dots, x_m\} \subset \mathbb{X}$ ,  $R_a$  with all  $d_i$  and empty set  $R_b$ , sorted list  $X$  with all  $x_i$  and empty list  $C$ ,  $G = (R_a \cup X_a, E)$  with  $(d_i, x_j) \in E$  if  $d_i \in N(x_j)$ , maximum iterations  $\text{maxdeep}$

**Ensure:** A minimum covers  $Z$  of  $R_a = \{d_1, \dots, d_n\}$  with elements in  $\mathbb{X}$ .

```

for every  $x_i, x_j \in X$  do
2:   if  $N(x_i) = N(x_j)$  then
        $x_i = \{x_i \text{ OR } x_j\}$ , remove  $x_j$ 
4:   end if
end for
6:  $k \leftarrow 0$ 
while  $|X| > 0$  AND  $k \leq \text{maxdeep}$  do
8:   for every  $d \in R_a$  do
       choose  $x_j \in N(d_i)$  with  $\max |N(x_j)|$ 
10:  for every  $d \in N(x_j)$  do
        $R_b \leftarrow d$ , from  $R_a$ .remove( $d$ )
12:  end for
       move  $x_j$  to  $Z_k$ 
14:  end for
        $R_a = R_b$ ,  $R_b = \emptyset$ ,  $k = k + 1$ 
16: end while
if  $R_a \neq \emptyset$  then
18:  add  $x_\infty$  to last  $Z_k$ 
end if
20: return  $Z = \min_{i \in \{1, \dots, k\}} Z_i$ ,
    
```

---

error  $Z = \min_{e(x_i)} Z_i$ , that means for  $q(Z) = C \setminus R$  is minimal. The pre-sorting of the context data list  $X$  results in covers of ascending cardinality, so the number of iterations  $k$  may be a limit for maximum cardinality. The pre-sorting can be removed, which results in more balanced and random

covers, whereof one with minimum error can be chosen.

## V. EXPERIMENTAL RESULTS

We tested our novel approach within two scenarios. First of all, using an artificial random instances with  $|\mathbb{D}| = 150$  documents and a given subset  $R$  with 20 example documents. We created instances with a fixed number of 80 or 40 normal distributed keywords which had a significant impact on the output. In addition we used  $N$  iterations, which lead to a different precision. The second scenario is a real-world example using set  $R$  of 10 random documents out of a human curated topic. We tested against complete PubMed Database using SCAIView. Thus  $|\mathbb{D}| \approx 29,000,000$ .

Within the random instances we were unable to describe a single document by its random keywords. This approach usually returned more than 100 documents. The reason for this rather contradictory result is still not entirely clear, but the normal distribution of keywords may be responsible for this result. The algorithms Tree-Description and Hierarchical Bipartite Cover-Description performed quite well, see figure 5. In general, we found Hierarchical Bipartite Cover-Description to work better and faster.

Changing to the real-world scenario the situation changes significantly. Given a set of 10 documents, Hierarchical Bipartite Cover-Description usually returned more than 6,000,000 documents, Tree-Description more than 5,000,000 before reaching the search-query length limitations. Vice versa we found, that the combination of keywords described a single document very well – even within nearly 3 million documents in  $\mathbb{D}$ . The keywords using MeSH-terms in PubMed are manually curated and seem not to be normally distributed.

The output of Keyword-Cover for 10 random examples with  $|R| = 10$  is presented in figures 6 and 7. The precision was iterated from 0.9 to 0.4. The output scales very well and is quite stable till precision 0.5 where we found between 12 and 36 documents. For precision 0.4 we found 28 till 676 documents.

We can see, that we have found a novel solution for search query finding on literature that performs quite well on real-world data. Our work clearly has some limitations. It is not clear, why the proposed algorithms perform significantly different in both scenarios. Despite this we believe our work could be the basis for solving the SQF and TD. Further work needs to be performed to the distribution of descriptive elements to documents to establish whether they can be used to generate search queries and topic descriptions that are significant enough.

## VI. CONCLUSIONS

We presented a novel formulation of both search query and topic finding problems as Minimum Set-Cover Problems. We proposed a weighted and unweighted version of the Minimum Document-Cover Problem as well as a hierarchical version using both AND as well as OR and the non-hierarchical version only using and.

With this we get a solution that uses on the one hand as much descriptive elements as possible to get as less documents in  $\mathbb{D}$  but not in  $R$ .

The search queries are not human readable. For example the tree-approach returns queries in the form `MeSH_Terms: D000818" AND ( "MeSH_Terms: D051381" OR "MeSH_Terms: D009538" OR "MeSH_Terms: D017207" OR "MeSH_Terms: Q000494" OR "MeSH_Terms: D006624" OR "MeSH_Terms: D011978" OR "MeSH_Terms: D000109" OR "MeSH_Terms: D008297" OR "MeSH_Terms: Q000187" OR "MeSH_Terms: Q000502" OR "MeSH_Terms: Q000378" OR "MeSH_Terms: D008464" OR "MeSH_Terms: Q000187" OR "MeSH_Terms: Q000187" OR ...`. This can be easily translated into something human-readable. But still it is a good probability that further research has to be done on how to shorten this to be both precise as well as significant.

In general this is both: a correct solution of clustering labeling of  $R$  on  $\mathbb{X}$  obtained by  $f$  as well as a possible solution of a search query so that  $q(Z) = R$ . It is not necessary an optimal solution of SQF or CLF problem, since reordering the keywords may result in better solutions.

The bipartite graph algorithms can be modified for many different use cases. All hierarchical algorithms can also be modified by adding weights. As described, there are many possible variations like sorting the context data list by minimum or maximum degree. The number of iterations  $k$  also has a big impact on the result. Another possible optimization is the pre-sorting by weighting the  $x_i$  with maximum  $|N(x_i)|$  and minimal  $D \setminus R$ .

This paper has underlined the importance of finding the computational core of NLP problems. We have managed to find a Minimum Set-Cover reformulation of SQF and TF which lead to an accurate solving of both on real-world data. The current study was unable to reproduce this success on random input data. Thus it is recommend that further research should be undertaken to examine the impact of keyword (or descriptive elements) distributions on documents. Nevertheless these results have been very encouraging to integrate this feature in SCAIView and to do further research on the optimization and extension of this heuristic.

## REFERENCES

- [1] J. Dörpinghaus, J. Darms, and M. Jacobs, "What was the question? a systematization of information retrieval and nlp problems." in *2018 Federated Conference on Computer Science and Information Systems (FedCSIS)*. IEEE, 2018.
- [2] N. R. Coordinators, "Database resources of the national center for biotechnology information," *Nucleic acids research*, vol. 45, no. Database issue, p. D12, 2017.
- [3] D. Suryanarayana, S. M. Hussain, P. Kanakam, and S. Gupta, "Natural language query to formal syntax for querying semantic web documents," in *Progress in Advanced Computing and Intelligent Engineering*. Springer, 2018, pp. 631–637.
- [4] D. Melo, I. P. Rodrigues, and V. B. Nogueira, "Semantic web search through natural language dialogues," in *Innovations, Developments, and Applications of Semantic Web and Information Systems*. IGI Global, 2018, pp. 329–349.

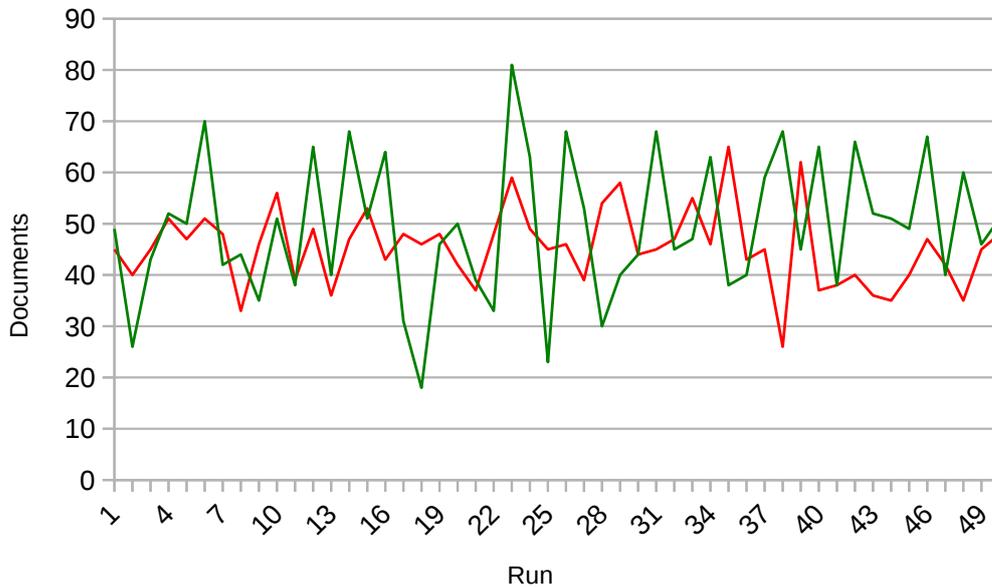


Fig. 5: Output of 50 random example runs and the number of retrieved documents in the artificial random scenario for algorithms Tree-Description (green) and Hierarchical Bipartite Cover-Description (red). The total number of documents was 150, and the document subset contains 20 documents. The number of keywords was 40. The number of iterations is  $N = 4$ .

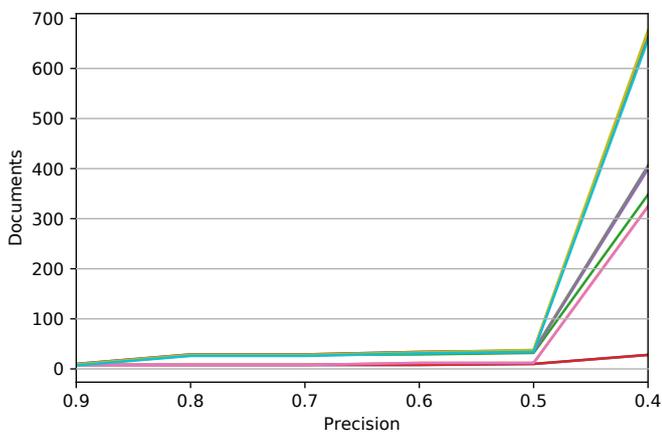


Fig. 6: Output of 10 random example runs with  $|R| = 10$  on PubMed. The precision was iterated from 0.9 to 0.4. The output scales very well and is quite stable till precision 0.5.

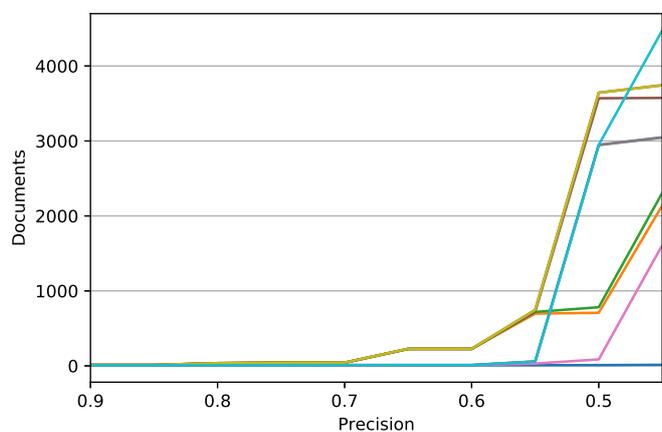


Fig. 7: Output of 10 random example runs with  $|R| = 10$  on PubMed. The precision was iterated from 0.9 to 0.4. The output scales very well and is quite stable till precision 0.5.

[5] J. Lin and W. J. Wilbur, “Pubmed related articles: a probabilistic topic-based model for content similarity,” *BMC bioinformatics*, vol. 8, no. 1, p. 423, 2007.  
 [6] D. Newman, S. Karimi, and L. Cavedon, “Using topic models to interpret medline’s medical subject headings,” in *Australasian Joint Conference on Artificial Intelligence*. Springer, 2009, pp. 270–279.  
 [7] D. Trieschnigg, P. Pezik, V. Lee, F. De Jong, W. Kraaij, and D. Rebholz-

Schuhmann, “Mesh up: effective mesh text classification for improved document retrieval,” *Bioinformatics*, vol. 25, no. 11, pp. 1412–1418, 2009.  
 [8] Z. Lu, W. J. Wilbur, J. R. McEntyre, A. Iskhakov, and L. Szilagy, “Finding query suggestions for pubmed,” in *AMIA Annual Symposium Proceedings*, vol. 2009. American Medical Informatics Association, 2009, p. 396.

- [9] A. A. Bertossi, "Dominating sets for split and bipartite graphs," *Information processing letters*, vol. 19, no. 1, pp. 37–40, 1984.
- [10] M. Yannakakis and F. Gavril, "Edge dominating sets in graphs," *SIAM Journal on Applied Mathematics*, vol. 38, no. 3, pp. 364–372, 1980.
- [11] B. Korte, J. Vygen, B. Korte, and J. Vygen, *Combinatorial optimization*. Springer, 2012, vol. 2.
- [12] P. Camerini, G. Galbiati, and F. Maffioli, "Complexity of spanning tree problems: Part i," *European Journal of Operational Research*, vol. 5, no. 5, pp. 346 – 352, 1980. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/0377221780901642>
- [13] E. Balas and M. W. Padberg, "On the set-covering problem," *Operations Research*, vol. 20, no. 6, pp. 1152–1161, 1972.
- [14] V. V. Vazirani, *Approximation algorithms*. Springer Science & Business Media, 2013.

# KPIs for Optimal Location of charging stations for Electric Vehicles: the Biella case-study

Edoardo Fadda, Daniele Manerba, Gianpiero Cabodi, Paolo Camurati, Roberto Tadei

Department of Control and Computer Engineering, Politecnico di Torino

Corso Duca degli Abruzzi, 24 - I-10129 Turin, Italy.

Email: {edoardo.fadda, daniele.manerba, gianpiero.cabodi, paolo.camurati, roberto.tadei}@polito.it

**Abstract**—Electric vehicles are accelerating the world’s transition to sustainable energy. Nevertheless, the lack of a proper charging station infrastructure in many real implementations still represents an obstacle for the spread of such a technology. In this paper, we present a real case application of optimization techniques in order to solve the location problem of electric charging stations in the district of Biella, Italy. The plan is composed by several progressive installations and decision makers pursue several objectives that might be in contrast. For this reason, we present an innovative framework based on the comparison of several ad-hoc Key Performance Indicators for evaluating many different aspects of a location solution.

## I. INTRODUCTION

Environmental pollution is one of the biggest problems affecting human society, and one of the main source of pollution is represented by motorized vehicles. It has been estimated that they are responsible for 40% of carbon dioxide emissions and 70% of other GHG emissions in urban areas [1]. In order to reduce this kind of pollution, an alternative and promising mobility solution is represented by the adoption of electric vehicles (EVs). Nevertheless, the expansion of this technology is strictly linked with the growth of a proper infrastructure for recharging the vehicles.

In this context, the company *Ener.bit S.r.l.*<sup>1</sup> and the *Department of Control and Computer Engineering* of Politecnico di Torino have recently developed a project for the sustainability of electric mobility in the district of Biella, Piedmont (Italy). The project goal was to plan the type, number, and location of the charging stations over an horizon of about 10 years (2019-2030). It is worthwhile noticing that the number of stations to locate and the number of power plugs for each station depend on an economical analysis related to the forecast number of EVs. Instead, the type of charging stations mainly depends on the features of a selected location. For example, a charging station near working centers can have a low charging system (because workers are assumed to park their vehicle during the entire day), while a charging station near shopping centers is supposed to be faster (cars must be recharged during shopping time). Therefore, the actual decision problem faced in the

project was to select the municipalities in the Biella district where to locate at least one charging station.

In general, location problems consider several different (and possibly conflicting) objectives, e.g., achieving a level of service proportional to the importance of the location, reducing the worst-case service level, maximizing the average service level, etc. Considering all those objectives in the same mathematical problem may end up with a huge amount of solutions that can confuse the decision maker instead of providing help. For this reason, our study provides an innovative analysis based on the comparison of several different aspects of a location solution through the use of a battery of Key Performance Indicators (KPIs). Moreover, since charging infrastructures are commonly supposed to be located through several progressive interventions over a defined time-horizon, we also analyze the trend of the provided KPIs over the interventions to generate long-term managerial insights.

### A. Literature review

Optimal location is a standard topic in operations research. There is a huge amount of different models, and the choice of the most correct model to abstract the problem depends on the objectives set and the constraints imposed by the application itself. In our case, it is fundamental to provide a constraint on the exact number of municipalities where to locate a charging station. Furthermore, the model should aim at optimizing some quality-of-service metrics for the user community.

In the literature, several works are present in this context. In [1], the authors present a study on the location of charging stations for EVs for the city of Lisbon (Portugal), characterized by a strong concentration of population and movements. The methodology is based on a model that maximizes demand coverage while maintaining an acceptable level of service. In [2], instead, the authors use a bilevel model in order to optimize vehicle sharing systems.

After a careful study of the existing approaches, and considering the specific features of the application at hand and the requests by the involved company, we decided to analyze the *p-centdian* model, which represents a combination of the classical *p-center* and *p-median* problems [3].

The rest of this paper is organized as follows. Section II is devoted to present the location model used in the project. In Section III, we propose and discuss several different KPIs of

This work has been supported by Ener.bit S.r.l. (Biella, Italy) under the grants "Studio di fattibilità per la realizzazione di una rete per la mobilità elettrica nella provincia di Biella" and "Analisi per la realizzazione di una rete per la mobilità elettrica nella provincia di Biella".

<sup>1</sup>Official website: <http://www.enerbit.it/>, last accessed: 2019-04-30.

interest for our application. In Section IV, we describe more in details the project and we present the numerical results. Finally, conclusions are drawn in Section V.

## II. THE P-CENTDIAN MODEL

Throughout the paper we use the following notation:

- $G = (N, E)$ : complete undirected graph with a set of nodes  $N$  representing possible locations for the charging stations and a set of edges  $E = \{(i, j) | i, j \in N, i \leq j\}$ ;
- $d_{ij}$ : distance between node  $i$  and node  $j \in N$  (note that distance  $d_{ii}$  may be non-null since it represents the internal distance to travel within municipality  $i \in N$ );
- $Q_i$ : service demand in node  $i \in N$ ;
- $h_i = Q_i / \sum_{j \in N} Q_j$ : demand rate of node  $i \in N$ ;
- $p$ : predefined number of stations to locate, with  $p \leq |N|$ ;
- $\bar{d}$ : coverage radius, i.e. the threshold distance to discriminate the covering. It represents, e.g., the maximum distance that an EV can travel (due to the battery capacity) or that a user is willing to drive to reach a charging station;
- $\mathcal{C}_i = \{j \in N, d_{ij} \leq \bar{d}\}$ : covering set of  $i \in N$ , i.e. the set of all stations nearer than  $\bar{d}$  from node  $i$ .

The *p-centdian* problem is to find  $p$  nodes where to locate charging stations so as to minimize a linear combination among the maximum and the average (weighted) distance between the located stations and the demand nodes. Its formulation is:

$$\min \lambda M + (1 - \lambda) \sum_{i \in N} h_i \sum_{j \in N | (i,j) \in E} d_{ij} x_{ij} \quad (1)$$

subject to

$$M \geq \sum_{j \in N | (i,j) \in E} h_i d_{ij} x_{ij} \quad \forall i \in N \quad (2)$$

$$\sum_{j \in N | (i,j) \in E} x_{ij} = 1 \quad \forall i \in N \quad (3)$$

$$\sum_{j \in N} y_j = p \quad (4)$$

$$\sum_{i \in N | (i,j) \in E} x_{ij} \leq |N| y_j \quad \forall j \in N \quad (5)$$

where  $y_j$  is a binary variable taking value 1 iff a station is located in node  $j \in N$ , and 0 otherwise, while  $x_{ij}$  is a binary variable taking value 1 iff the demand of node  $i \in N$  is served by a charging station located in  $j \in N$ , and 0 otherwise.

The objective function (1) consists of a linear combination of two terms. The first is the auxiliary variable  $M$  that, according to constraints (2), takes the maximum value of the expression  $\sum_{j \in N} h_i d_{ij} x_{ij}$  over all nodes  $i \in N$ . In other words, it is the maximum distance between a demand node and its closest station. The second is the average distance traveled by the total demand flow towards charging stations. Clearly, through the parameter  $0 \leq \lambda \leq 1$  it is possible to define the relative importance of one objective with respect to the other one. In this work, we set the  $\lambda$  parameter dynamically by using

the ratio between the optima of the relative *p*-center and *p*-median subproblems. In this way we ensure that the two terms of (1) are comparable. Constraints (3) ensure that each demand node is served by exactly one station. Constraint (4) ensures to locate exactly  $p$  stations. Finally, logical constraints (5) ensure to locate a station in  $j$  (i.e.,  $y_j = 1$ ) only if it is assigned to serve at least one demand node (i.e.,  $\sum_{i \in N} x_{ij} > 0$ ).

## III. KEY PERFORMANCE INDICATORS

In this section, we define the set of KPIs that were used in the project in order to measure the performance of the solution provided by the model. For simplicity, we define  $\mathcal{L}_i = \{j \in \mathcal{C}_i | y_j = 1\}$  as the set of nodes where a charging station has been located that covers demand node  $i$ , and  $\mathcal{C} = \{i \in N | \exists j \in \mathcal{C}_i \text{ such that } y_j = 1\}$  as the set of demand nodes covered by at least one charging station.

The following proposed KPIs consider topological, coverage, and accessibility measures:

- **WORST-CASE DISTANCE:**

$$D_{max} := \max_{i \in N} \min_{j \in \mathcal{L}} d_{ij} \quad (6)$$

represents the maximum distance between a demand node and its closest charging station.

- **BEST-CASE DISTANCE:**

$$D_{min} := \min_{i \in N} \min_{j \in \mathcal{L}} d_{ij} \quad (7)$$

represents the minimum distance between a demand node and its closest charging station.

- **AVERAGE DISTANCE:**

$$D_{avg} := \frac{1}{|N|} \sum_{i \in N} \min_{j \in \mathcal{L}} d_{ij} \quad (8)$$

represents the average distance between a demand node and its closest charging station.

- **DISPERSION:**

$$Disp := \sum_{i \in \mathcal{L}} \sum_{j \in \mathcal{L}} d_{ij} \quad (9)$$

represents the sum of the distances between all the located stations. It is a measure of homogeneity of the service from a purely topological point of view.

- **ACCESSIBILITY:**

$$Acc := \sum_{i \in N} h_i A_i \quad (10)$$

is the total accessibility of the charging service, where

$$A_i := \sum_{j \in \mathcal{L}} e^{-\beta d_{ij}} \quad (11)$$

is the accessibility of a facility in the sense of [4]. The parameter  $\beta > 0$  must be calibrated and represents the dispersion of the alternatives in the choice process (the calibration has been performed according to [5] and [6]).

- **COVERAGE:**

$$C := 100 * |\mathcal{C}| / |N| \quad (12)$$

represents, in percentage, the number of covered locations with respect to the total.

- WORST-CASE COVERAGE:

$$C_{min} := \min_{i \in N} |\mathcal{L}_i| \quad (13)$$

represents the minimum number of charging stations covering a demand node.

- BEST-CASE COVERAGE:

$$C_{max} := \max_{i \in N} |\mathcal{L}_i| \quad (14)$$

represents the maximum number of charging stations covering a demand node.

- AVERAGE COVERAGE:

$$C_{avg} := \frac{1}{N} \sum_{i \in N} |\mathcal{L}_i| \quad (15)$$

represents the average number of charging stations covering a demand node.

#### IV. THE BIELLA CASE-STUDY

In the aforementioned project, the possible locations are the 78 municipalities of the district of Biella, Italy. From a preceding economical analysis, the company is supposed to install charging stations in one municipality by the end of 2019, in 10 municipalities by the end of 2022, in 37 by the end of 2025, and in all remaining municipalities by the end of 2030. Moreover, the company assumed a coverage radius  $\bar{d} = 25$ , i.e., a municipality is covered if its distance from the nearest charging station is less than 25 kilometers. We remark that each station may have different size, number of plugs, and capacity in terms of charging. However, as already stated in the Introduction, we just focus on selecting the municipalities of Biella district where to locate at least one charging station, while the real characteristics of the stations will be derived in a successive phase. For example, the number of plugs for each municipality can be calculated as a proportion to the demand rate of that particular municipality (and its surroundings).

The p-centdian model, accurately instantiated with the data deriving from the Biella district case study, can be easily solved by exact algorithms as the branch-and-cut implemented in the available commercial and academic solvers. In our particular case, we used the GUROBI solver v.8.1.0. The resolution was performed on a common PC (Intel Core i7-5500U CPU@2.40 GHz with 8 GB RAM) and took on average 12 seconds. Notice how the resolution efficiency obtained allows to possibly perform a large number of experiments with different input data, thus refining the analysis.

The solutions for the different time thresholds studied, obtained using the p-centdian model, are the following (clearly, at each intervention, the locations chosen in the previous steps are forced to remain in the solution):

- one municipality ( $p = 1$ ) by the end of 2019: the only municipality chosen is Biella, the chief town (see Figure 1). This was expected since Biella is the most important city in terms of demand.

- 10 municipalities ( $p = 10$ ) by the end of 2022: some small municipalities close to and other big ones far from Biella are chosen (see Figure 2).
- 37 municipalities ( $p = 37$ ) by the end of 2025: the solution tends to select municipalities close to the previously selected ones, creating clusters (see Figure 3)
- all municipalities ( $p = 78$ ) by the end of 2030 (this corresponds to the trivial solution with  $y_i = 1, \forall i \in N$ ).

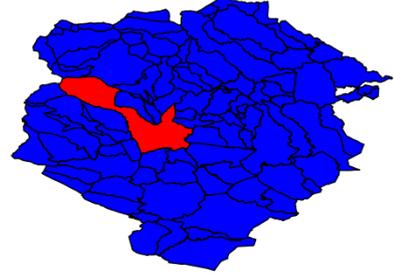


Fig. 1. Optimal location for  $p = 1$  (2019). Chosen locations in red.

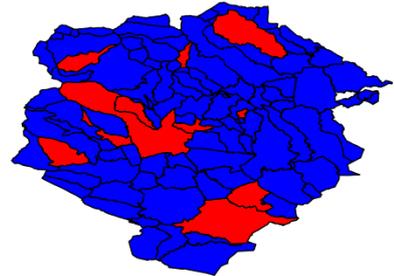


Fig. 2. Optimal location for  $p = 10$  (2022). Chosen locations in red.

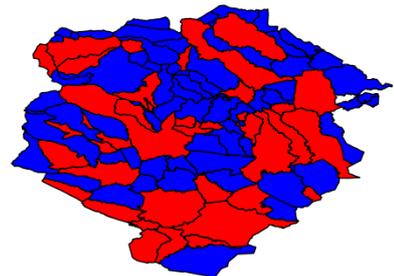


Fig. 3. Optimal location for  $p = 37$  (2025). Chosen locations in red.

The value of all the KPIs, in the various steps of intervention, is calculated and shown in Table I. Note that the last column, corresponding to the case in which all the locations are chosen, contains the best possible value for each KPI. Several observations can be done:

- $D_{max}$  decreases with the increase in the number of municipalities in which at least one charging station has been located and, as it can be seen, it reaches reasonable values from  $p = 10$  onward.
- $D_{min}$  decreases as the number of municipalities in which at least one charging station has been located increases, and it stabilizes at the best value already with  $p = 10$ .
- $D_{avg}$  decreases as the number of municipalities in which at least one charging station has been located increases. It is interesting to note that the percentage improvement in the indicator decreases as the number of selected municipalities increases.
- $Disp$  increases as the number of municipalities in which at least one charging station has been located increases. Its growth is very marked due to the factorial growth of the number of pairs of selected municipalities. The starting value is set to zero since with a single municipality the summation in the definition cannot be calculated.
- $Acc$  increases as the number of municipalities in which at least one charging station has been located increases. Also in this case the improvements are less marked as the number of selected municipalities increases.
- $C$  increases as the number of municipalities in which at least one charging station has been located increases. It can be seen that with only 10 selected municipalities, the coverage reaches very high levels (96% of the municipalities are covered).
- $C_{min}$  increases with the number of municipalities where at least one charging station has been located. Since this is the most pessimistic case, this indicator remains at zero when 1, 10, and 37 selected municipalities are considered. The data then verifies the non-total coverage shown by the KPI previously discussed.
- $C_{max}$  increases as the number of municipalities in which at least one charging station has been located increases. It can be seen that the increase in value grows with the number of selected municipalities. However, it can be noted that already with 10 municipalities the most covered municipality has the choice between 7 charging stations within a 25 kilometers radius.
- $C_{avg}$  increases with the increase in the number of municipalities in which at least one charging station has been located and, as it can be seen, has a much lower value than the  $C_{max}$ . This implies a heterogeneous situation in terms of coverage of the various locations. In fact, we have a large number of municipalities covered by a few charging stations and a small number of municipalities covered by many charging stations. Since the towns that are not covered are those with a lower demand (i.e., with less electric vehicles) this feature is in line with the technical specifications of the problem.

A common trend of almost all the KPIs is that the second intervention is the one providing the highest proportional change with respect to the previous one (e.g.,  $C$  almost doubles its value for  $p = 10$  while it gains only few units for  $p = 37$

TABLE I  
KPIs VALUE IN THE FOUR INTERVENTIONS.

KPI	$p = 1$ (2019)	$p = 10$ (2022)	$p = 37$ (2025)	$p = 78$ (2030)
$D_{max}$	53	24	20	11
$D_{min}$	5.7	2	2	2
$D_{avg}$	20.3	8.9	5.8	4.4
$Disp$	0	2158.2	34663.9	167201.3
$Acc$	0.024769	0.115986	0.329689	0.456748
$C$	55%	96%	98%	100%
$C_{min}$	0	0	0	1
$C_{max}$	1	7	22	43
$C_{avg}$	0.089744	2.653846	8.833333	19.28205

and  $p = 78$ ). Interesting enough,  $D_{min}$  reaches its optimal value even for  $p = 10$ . This represents a very important insight for the company for two main reasons. First, it means that the users will perceive the biggest improvement in terms of service in relatively small amount of time (the first 3-5 years) and in response to a small effort in terms of installed stations. Second, it means that the last interventions, which are the ones affected by the most uncertainty (e.g., in terms of economical sustainability), are not very critical for the process overall quality.

## V. CONCLUSIONS

The implementation of the plan resulting from this study in the district of Biella still needs a detailed urban planning and electrical plant analysis to determine the physical points within the municipalities in which to locate the charging stations identified. However, the described methodologies represent the application of state-of-the-art technology in optimal location to real problems. It is worthwhile noting that the developed analysis can be applied to different location models and to a broader set of KPIs. This way the decision maker can evaluate different solutions and generate insights for the location problem at hand.

## REFERENCES

- [1] I. Frade, A. Ribeiro, G. Goncalves, and A. Pais Antunes, "Optimal location of charging stations for electric vehicles in a neighborhood in lisbon, portugal," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2252, pp. 91–98, 2011.
- [2] A. Quiliot and A. Sarbinowski, "Facility location models for vehicle sharing systems," *Proceedings of the 2016 Federated Conference on Computer Science and Information Systems*, vol. 8, pp. 605–608, 2016.
- [3] H. A. Eiselt and C. L. Sandblom, *Decision Analysis, Location Models, and Scheduling Problems*. Springer-Verlag Berlin Heidelberg, 2004.
- [4] W. Hansen, "How accessibility shapes land use," *Journal of the American Institute of Planners*, vol. 25, pp. 73–76, 1959.
- [5] R. Tadei, N. Ricciardi, and G. Perboli, "The stochastic p-median problem with unknown cost probability distribution," *Operations Research Letters*, vol. 37, pp. 135–141, 2009.
- [6] E. Fadda, L. Fotio Tiotsop, D. Manerba, and R. Tadei, "The stochastic multi-path traveling salesman problem with dependent random travel costs," *Transportation Science (submitted)*, 2019.

# A novel integer linear programming model for routing and spectrum assignment in optical networks

Youssef Hadhbi, Hervé Kerivin, Annegret Wagler

Université Clermont Auvergne,

LIMOS (UMR 6158 CNRS),

Clermont-Ferrand, France

Email: {yousseuf.hadhbi,herve.kerivin,annegret.wagler}@uca.fr

**Abstract**—The routing and spectrum assignment problem is an NP-hard problem that receives increasing attention during the last years. Existing integer linear programming models for the problem are either very complex and suffer from tractability issues or are simplified and incomplete so that they can optimize only some objective functions. The majority of models uses edge-path formulations where variables are associated with all possible routing paths so that the number of variables grows exponentially with the size of the instance. An alternative is to use edge-node formulations that allow to devise compact models where the number of variables grows only polynomially with the size of the instance. However, all known edge-node formulations are incomplete as their feasible region is a superset of all feasible solutions of the problem and can, thus, handle only some objective functions.

Our contribution is to provide the first complete edge-node formulation for the routing and spectrum assignment problem which leads to a tractable integer linear programming model. Indeed, computational results show that our complete model is competitive with incomplete models as we can solve instances of the RSA problem larger than instances known in the literature to optimality within reasonable time and w.r.t. several objective functions. We further devise some directions of future research.

## I. INTRODUCTION

TODAY'S communication networks are optical networks where light is used as communication medium between sender and receiver nodes. For over two decades, the Wavelength-Division Multiplexing (WDM) has been the most popular technology used in fiber-optic communication. WDM combines multiple wavelengths to simultaneously transport signals over a single optical fiber, but must select the wavelengths from a rather coarse fixed grid of frequencies specified by the United Nations agency ITU (International Telecommunication Union) and leads to inefficient use of spectral resources and bans allocating more than a single wavelength to a traffic demand.

In response to the sustained growth of data traffic volumes in communication networks, a new generation of optical networks, called flexgrid Elastic Optical Networks (EONs), has been introduced in the last few years to enhance the spectrum efficiency and enlarge the network capacity [7].

In EONs, the frequency spectrum of an optical fiber is divided into many narrow frequency slots of fixed spectrum width. Any sequence of consecutive slots can form a channel

that can be switched in the network nodes to create a lightpath (i.e., an optical connection represented by a route and a channel). EONs enable capacity gain by allocating minimum required bandwidth thanks to a finer spectrum granularity than in the traditional WDM networks.

However, the spectrum assignment in EONs leads to the *Routing and Spectrum Assignment (RSA) problem* that is much harder to handle in practice than its counterpart using Wavelength-Division Multiplexing. In fact, the RSA problem consists of two parts: the routing (to select for each traffic demand a path through the communication network) and the spectrum assignment (to assign for each demand an interval of consecutive frequency slots within the optical spectrum such that the intervals of lightpaths using a same edge in the network are disjoint), see e.g. [15] and Section II for details. Thereby, the following constraints need to be respected when dealing with the RSA problem:

- 1) *spectrum continuity*: the frequency slots allocated to a demand remain the same on all the links of a route;
- 2) *spectrum contiguity*: the frequency slots allocated to a demand must be contiguous;
- 3) *non-overlapping spectrum*: a frequency slot can be allocated to at most one demand.

The RSA problem is a generalization of the well-studied *Routing and Wavelength Assignment (RWA) problem* that is associated with a fixed grid of frequencies [3].

The former problem has started to receive a lot of attention over the last few years. It has been shown to be NP-hard [2], [18]. In fact, if for each demand the route is already known, the RSA problem reduces to the so-called *Spectrum Assignment (SA) problem* and only consists of determining the demands' channels. The SA problem has been shown to be NP-hard on paths [14] which makes the SA problem (and thus also the RSA problem) much harder than the RWA problem which is well-known to be polynomially solvable on paths, see e.g. [3].

To solve the RSA problem, various approaches have been studied in the literature, based on different Integer Linear Programming (ILP) models. Hereby, detailed models aiming at precisely describing all technological aspects of EONs and being able to handle various criteria for optimization typically suffer from tractability issues resulting from their

greater complexity such that the tendency is to use simplified or restricted models.

The majority of the existing models uses an *edge-path formulation* where for each demand, variables are associated either with all possible routing paths or with all possible light-paths for this demand. One characteristic of this formulation is, therefore, an exponential number of variables issued from the total number of all feasible paths between origin-destination pairs in the network, which grows exponentially with the size of the network.

To bypass the exponential number of variables, edge-path formulations with a precomputed subset of all possible paths per demand have been studied e.g. in [8], [9], [16], [19], see [19] for an overview. However, such formulations cannot guarantee optimality of the solutions in general (as only a precomputed subset of paths is considered and, thus, a restricted problem solved). In order to be able to find optimal solutions of the RSA problem w.r.t. any objective function with the help of an edge-path formulation, all possible paths have to be taken into account. As the explicit models are far too big for computation, it is in order to apply column-generation methods. However, computational results from e.g. [10], [11], [13] show that the size of the instances that can be solved that way is rather limited.

An alternative to edge-path formulations is to use *edge-node formulations* that lead to less intuitive models for the routing, but have the advantage that the number of variables grows only polynomially with the size of the instance. Despite this advantage, edge-node formulations are not yet well-studied. Only few authors made use of this type of model, as Cai et al. [1], Velasco et al. [16], Zotkiewicz et al. [19], and Jia et al. who used in [6] an edge-node formulation to treat a more general problem.

All three models from [1], [16], [19] are compact models as both the number of variables and constraints is polynomial in terms of the size of the instance. However, all three models are incomplete as their feasible region is a superset of all feasible solutions of the RSA problem and can, thus, handle only some objective functions (see Section IV for details).

Our contribution is to provide the first complete edge-node formulation for the RSA problem that precisely encodes the set of all feasible solutions and can, therefore, be used to optimize any chosen objective function. For that, we propose an appropriate combination of variables and constraints (partly using new variables and constraints), see Section III for details. Our model uses, as in [1], [16], [19], a polynomial number of variables, but an exponential number of constraints to ensure the exact encoding of feasible solutions. As we are able to separate the exponentially-sized families of constraints in polynomial time, our model is computationally tractable and, therefore, competitive with the compact but incomplete models from [1], [16], [19].

While Zotkiewicz et al. [19] do not give computational results, Velasco et al. [16] tested their formulation on a network topology of Spain with 35 edges (64 slots per edge) and 21 nodes with a very small number of 12 demands and

requested numbers of slots in  $\{1, 2, 4\}$ . The results show that Cplex version 12 could optimally solve the problem after 6 hours by minimizing the number of edges activated for the routing (which can be looked as a network design problem).

Cai et al. [1] tested their formulation on two small network topologies, one with 6 nodes and 9 links and the other with 10 nodes and 22 links, one demand between each pair of nodes in the network and requested numbers of slots in  $\{1, \dots, 3\}, \dots, \{1, \dots, 9\}$ . The results show that Gurobi 5.0 could optimally solve the problem after 1 hour by minimizing the max-slot position for the 6 nodes and 9 links topology (but did not report on time limits to solve the instances on the other network).

Our model allows us to solve instances of the RSA problem larger than the instances in [1], [16] to optimality within reasonable time w.r.t. several objective functions (see Section V for details).

The paper is organized as follows. In Section II, we describe in detail the input and the desired output of the RSA problem together with the studied objective functions. In Section III, we present our new edge-node formulation and compare it in Section IV with existing models from the literature [1], [16], [19]. In Section V, we report on computational results achieved with the help of our formulation. We close with some concluding remarks and future research.

## II. THE RSA PROBLEM

In this section, we formally define the RSA problem by describing in detail the input and the desired output of the RSA problem together with the studied objective functions.

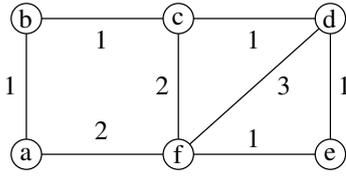
As input of the RSA problem, we are given

- an optical spectrum  $S = \{1, \dots, \bar{s}\}$  of available frequency slots;
- an optical network, represented as an undirected, loopless, connected graph  $G = (V, E)$  that may have parallel edges (if parallel optical fibers are installed between two nodes), and for each edge  $e \in E$  its length  $\ell_e \in \mathbb{R}_+$  (in kms),
- a multiset  $K$  of demands where each demand  $k \in K$  is specified by
  - an origin node  $o_k \in V$  and a destination node  $d_k \in V \setminus \{o_k\}$ ,
  - a requested number  $w_k \in \mathbb{N}_+$  of slots, and
  - a transmission reach  $\bar{\ell}_k \in \mathbb{R}_+$  (in kms).

The task is to determine for each demand  $k \in K$  a lightpath composed of an  $(o_k, d_k)$ -path  $P_k$  in  $G$  respecting the transmission reach  $\bar{\ell}_k$  and a subset  $S_k \subset S$  of  $w_k$  consecutive frequency slots that is available on all edges of  $P_k$  and disjoint from the subsets  $S_{k'}$  of all other demands  $k'$  routed along an edge of  $P_k$ , thereby minimizing some objective function.

Hence, the desired output of the RSA problem is, for each demand  $k \in K$ , a lightpath composed of

- an  $(o_k, d_k)$ -path  $P_k$  in  $G$  with  $\sum_{e \in E(P_k)} \ell_e \leq \bar{\ell}_k$ ,


 Fig. 1. The network  $G$  used in Example 2.1.

- a subset  $S_k \subset \{1, \dots, \bar{s}\}$  of  $w_k$  consecutive slots with  $S_k \cap S_{k'} = \emptyset$  for each demand  $k' \in K$  routed along an edge  $e \in E(P_k)$ .

This output can be given in terms of a matrix  $M \in \mathbb{N}^{|E| \times \bar{s}}$  with

$$M_{e,s} = \begin{cases} k & \text{if slot } s \in S \text{ is allocated to} \\ & \text{demand } k \in K \text{ on edge } e \in E, \\ 0 & \text{otherwise.} \end{cases}$$

In addition, the selected set of lightpaths is supposed to minimize a chosen objective function. In this paper, we will focus on the following objective functions that have been used in [1], [16], [19] to be able to compare our computational results with those from the literature:

- $O_1$ : minimize the sum of hops in paths (where the term hops refers to the number of edges in a path  $P_k$ ) [19],
- $O_2$ : minimize the number of edges from the network used to route the demands [16],
- $O_3$ : minimize the maximal used slot position (and, thus, the width of the subspectrum of  $S$  used for the spectrum assignment) [1].

Note that the first two objective functions are only related to the routing (provided that a feasible spectrum assignment within  $S$  exists for this routing), whereas the third objective function seeks for the most efficient spectrum assignment over all possible routings.

*Example 2.1:* Consider the following small instance of the RSA problem, given by a spectrum of width  $\bar{s} = 10$ , the network  $G$  shown in Figure 1 with edge length as indicated, and the following set  $K$  of demands:

$k$	$o_k \rightarrow d_k$	$w_k$	$\bar{l}_k$
1	$a \rightarrow c$	2	4
2	$a \rightarrow d$	1	4
3	$b \rightarrow f$	2	4
4	$b \rightarrow e$	1	4
5	$d \rightarrow f$	3	4

An optimal solution w.r.t. objective function

- $O_1$  with minimum sum 11 of hops in paths is represented by matrix  $M_1$ ,
- $O_2$  with minimum number 5 of edges from the network used to route the demands is represented by matrix  $M_2$ ,
- $O_3$  with minimum maximal used slot position 4 is represented by matrix  $M_3$ .

$$M_1 = \begin{pmatrix} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ ab & & & 3 & 3 & & 4 & & & & \\ af & 1 & 1 & 3 & 3 & 2 & 4 & & & & \\ bc & & & & & & & & & & \\ cd & & & & & & & & & & \\ cf & 1 & 1 & & & & & & & & \\ de & & & & & & 2 & & & & \\ df & 5 & 5 & 5 & & & & & & & \\ ef & & & & & & 2 & 4 & & & \end{pmatrix}$$

$$M_2 = \begin{pmatrix} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ ab & & & 3 & 3 & & 4 & & & & \\ af & 1 & 1 & 3 & 3 & 2 & 4 & & & & \\ bc & & & & & & & & & & \\ cd & & & & & & & & & & \\ cf & 1 & 1 & & & & & & & & \\ de & 5 & 5 & 5 & & & 2 & & & & \\ df & & & & & & & & & & \\ ef & 5 & 5 & 5 & & & 2 & 4 & & & \end{pmatrix}$$

$$M_3 = \begin{pmatrix} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ ab & & & & 4 & 2 & & & & & \\ af & 1 & 1 & 4 & & & & & & & \\ bc & 3 & 3 & & 2 & & & & & & \\ cd & 3 & 3 & & 2 & & & & & & \\ cf & 1 & 1 & & & & & & & & \\ de & 3 & 3 & & & & & & & & \\ df & 5 & 5 & 5 & & & & & & & \\ ef & 3 & 3 & 4 & & & & & & & \end{pmatrix}$$

### III. A NOVEL EDGE-NODE FORMULATION

In this section we introduce our novel edge-node ILP model for the RSA problem in the general variant where demands may be rejected.

a) *Variables:* For the routing, *demand-edge variables*

$$x_e^k = \begin{cases} 1 & \text{if demand } k \text{ is routed through edge } e, \\ 0 & \text{otherwise,} \end{cases}$$

are used for all  $k \in K$  and all  $e \in E$  as in [17], [19].

For the spectrum assignment, several different variables are necessary. As in [2], [16], *demand-slot variables*

$$z_s^k = \begin{cases} 1 & \text{if slot } s \text{ is the last slot allocated for demand } k, \\ 0 & \text{otherwise,} \end{cases}$$

are used which indicate that  $s$  is the last of the  $w_k$  consecutive slots allocated for the demand  $k \in K$ , with  $s \in S$ . The consecutive slots  $s' \in \{s - w_k + 1, \dots, s\}$  shall form the channel assigned to this demand  $k$  whenever  $z_s^k = 1$ .

We newly propose *demand-edge-slot variables*

$$t_{e,s}^k = \begin{cases} 1 & \text{if slot } s \text{ is assigned to demand } k \text{ on edge } e, \\ 0 & \text{otherwise,} \end{cases}$$

for all demands  $k \in K$ , all edges  $e \in E$  and all slots  $s \in S$ .

When we optimize objective functions involving max-used slot positions, we newly propose *edge-max-slot-position variables*  $p_e \in \mathbb{Z}^+$  for all edges  $e \in E$  (which indicate the position

of the last slot allocated on the edge  $e \in E$ ), as well as a *max-slot-position variable*  $p \in \mathbb{Z}^+$  (which represents the position of the highest slot used over all the edges  $e \in E$  as in [2]).

When we optimize the number of edges used for the routing, we newly propose *edge-activation variables*

$$a_e = \begin{cases} 1 & \text{if some demand } k \text{ is routed through edge } e, \\ 0 & \text{otherwise,} \end{cases}$$

for all edges  $e \in E$ .

*b) Constraints:* To formulate the constraints, we employ the following notations. For any non-empty subset  $X \subset V$ , let  $\delta(X)$  denote the set of edges having one endnode in  $X$  and the other endnode in  $V \setminus X$ . The pair  $(X, V \setminus X)$  is called a cut of  $G$ , the edges in  $\delta(X)$  are said to cross this cut. In the special case  $X = \{v\}$ , we write  $\delta(v)$  instead of  $\delta(\{v\})$ .

For the routing, we use demand-edge variables  $x_e^k$  and have to ensure by appropriate constraints that the subset

$$E(k) = \{e \in E : x_e^k = 1\}$$

of edges selected for the routing of demand  $k$  indeed forms an  $(o_k, d_k)$ -path  $P_k$  in  $G$ , for each demand  $k \in K$ . For that, we use the following constraints. The *origin constraints*

$$\sum_{e \in \delta(o_k)} x_e^k \leq 1, \text{ for all } k \in K \quad (1)$$

ensure that at most one path  $P_k$  can leave the origin  $o_k$  as at most one of the edges  $e \in \delta(o_k)$  incident to  $o_k$  can be selected for  $E(k)$ . Similarly, *destination constraints*

$$\sum_{e \in \delta(d_k)} x_e^k - \sum_{e \in \delta(o_k)} x_e^k = 0, \text{ for all } k \in K \quad (2)$$

force that the path  $P_k$  enters its destination  $d_k$ , provided that there is a path  $P_k$  leaving  $o_k$ . (Note that if no path is selected for demand  $k$ , then  $\sum_{e \in \delta(o_k)} x_e^k = 0$  holds and ensures that no edge from  $\delta(d_k)$  can be selected either for  $E(k)$ .) Origin and destination constraints are used in [1], [16], [19] in a slightly different manner.

In addition, we newly propose *path-continuity constraints*

$$\sum_{e \in \delta(X)} x_e^k - \sum_{e \in \delta(o_k)} x_e^k \geq 0, \forall k \in K, \forall X, o_k \in X, d_k \in V \setminus X. \quad (3)$$

These constraints are important whenever a path  $P_k$  is selected for demand  $k$  (and, thus,  $\sum_{e \in \delta(o_k)} x_e^k = 1$  holds): they guarantee that there is an edge  $e \in \delta(X) \cap E(k)$  such that the path  $P_k$  indeed crosses the cut  $(X, V \setminus X)$  for each  $X$  with  $o_k \in X$  and  $d_k \in V \setminus X$ .

Hence, origin, destination and path-continuity constraints together imply that  $E(k)$  contains an  $(o_k, d_k)$ -path  $P_k$ . It is left to prevent  $E(k)$  from having more edges than needed for  $P_k$  and  $P_k$  from having a length exceeding the transmission reach of demand  $k$ .

For that, we use as in [6], [16] *degree constraints*

$$\sum_{e \in \delta(v)} x_e^k \leq 2, \text{ for all } k \in K, \text{ and all } v \in V \setminus \{o_k, d_k\} \quad (4)$$

to prevent that more than two edges from  $E(k)$  are incident to any node. Furthermore, we newly propose *cycle-elimination constraints*

$$\sum_{e' \in \delta(X_e)} x_{e'}^k \geq \begin{cases} 2x_e^k & \text{if } |X_e \cap \{o_k, d_k\}| = 0 \\ x_e^k & \text{if } |X_e \cap \{o_k, d_k\}| = 1 \end{cases} \quad (5)$$

$$\forall k \in K, \forall e \in E, \forall X_e \subset V$$

where  $X_e \subset V$  denotes a subset of nodes containing both endnodes of edge  $e$ , to avoid cycles isolated from  $P_k$  (note that isolated edges also fall into this case).

Moreover, we newly propose a *transmission-reach constraint*

$$\sum_{e \in E} l_e x_e^k - \bar{l}_k \sum_{e \in \delta(o_k)} x_e^k \leq 0, \text{ for all } k \in K \quad (6)$$

to ensure that the length of  $P_k$  does not exceed the transmission reach of  $k$  if the demand  $k$  is accepted, otherwise all the variables  $x_e^k$  are forced to equal zero.

When we optimize the number of edges used for the routing, we need in addition the following constraints

$$a_e - x_e^k \geq 0, \text{ for all } k \in K, \text{ and all } e \in E \quad (7)$$

to force  $a_e = 1$  when  $x_e^k = 1$  for some  $k \in K$ , and

$$a_e \leq \sum_{k \in K} x_e^k, \text{ for all } e \in E \quad (8)$$

to guarantee  $a_e = 0$  if edge  $e$  is not used in any routing.

For the spectrum assignment, we have to guarantee that, whenever demand  $k$  is accepted and an  $(o_k, d_k)$ -path  $P_k$  has been selected,

- a channel  $S_k \subset S$  of  $w_k$  consecutive frequency slots is assigned to  $k$ ,
- this channel is the same on all edges of  $P_k$  and disjoint from the channels  $S_{k'}$  of all other demands  $k'$  routed along an edge of  $P_k$ .

We newly propose *channel selection constraints*

$$\sum_{s=w_k}^{\bar{s}} z_s^k - \sum_{e \in \delta(o_k)} x_e^k = 0, \text{ for all } k \in K \quad (9)$$

that do not allow to assign a channel to demand  $k$  when no path  $P_k$  is selected (by not allowing to assign a slot  $s$  as last slot in the channel), but force to select such a last slot in the channel whenever a path is leaving  $o_k$ . In addition, we specify the available last slots for the channel of demand  $k$  by *forbidden-slot constraints*

$$\sum_{s=1}^{w_k-1} z_s^k = 0, \text{ for all } k \in K, \quad (10)$$

to prevent demand  $k$  to occupy a slot  $s$  as last slot in the channel whenever  $s < w_k$ . Klinkowski et al. [9] proposed a similar idea using demand-edge-first-slot variables.

We newly propose *edge-slot constraints*

$$\sum_{s \in S} t_{e,s}^k - w_k x_e^k = 0, \text{ for all } k \in K \text{ and all } e \in E \quad (11)$$

to ensure that precisely  $w_k$  slots are allocated on edge  $e$  to demand  $k$  if and only if demand  $k$  is routed through edge  $e$ .

Spectrum contiguity and continuity are handled by the following new *demand-edge-slot constraints*

$$x_e^k + \sum_{s' = s}^{\min(s+w_k-1, \bar{s})} z_{s'}^k - t_{e,s}^k \leq 1, \forall k \in K, \forall e \in E, \forall s \in S \quad (12)$$

to force that slot  $s$  on edge  $e$  is allocated to demand  $k$  if and only if demand  $k$  passes through edge  $e$  and slot  $s$  belongs to the channel assigned to demand  $k$  (which is the case if one slot  $s' \in \{s, \dots, s + w_k - 1\}$  is the last slot of the channel).

We newly propose *non-overlapping constraints*

$$\sum_{k \in K} t_{e,s}^k \leq 1, \text{ for all } e \in E \text{ and all } s \in S \quad (13)$$

to ensure that a slot  $s$  on edge  $e$  can be allocated to at most one demand.

When we optimize objective functions involving max-used slot positions, we newly propose two additional constraints

$$st_{e,s}^k - p_e \leq 0, \text{ for all } k \in K, \text{ all } e \in E \text{ and all } s \in S \quad (14)$$

to guarantee that no slot  $s$  above  $p_e$  is used on edge  $e$  and

$$p_e - \sum_{k \in K} \sum_{s \in S} st_{e,s}^k \leq 0, \text{ for all } e \in E \quad (15)$$

to force the max used slot position on edge  $e$  to equal 0 if no demand is routed through edge  $e$ , set the bounds  $p_e \leq p \leq \bar{s}$ , and force  $p_e \in \mathbb{N}$  for all  $e \in E$  and  $p \in \mathbb{N}$  to be integral. Finally, we force all other variables to be binary and require non-negativity for all variables.

c) *Objective functions*: With the help of these variables, the considered objective functions read as follows:

- $\min \sum_{e \in E, k \in K} x_e^k$  to minimize the sum of number of hops in the paths,
- $\min \sum_{e \in E} a_e$  to minimize the number of edges used for the routing, and
- $\min p$  to minimize the max-used slot position.

Recall that our model encodes the general variant of the RSA problem when demands may be rejected. This situation does not comply with the objective functions studied in [1], [16], [19] (as for all three objective functions, rejecting all demands would yield the optimal solution, with objective function value equal to 0). Our model can be easily adapted to the special case where all demands have to be served, by requiring equality in the origin constraint (1) and simplifying the constraints (2), (3), (6) and (9) by replacing the term  $\sum_{e \in \delta(o_k)} x_e^k$  by 1.

#### IV. COMPARISON OF EDGE-NODE FORMULATIONS

All three edge-node formulations from [1], [16], [19] for the RSA problem are compact models as both the numbers of variables and constraints grow only polynomially in the size of the instance, i.e., in the size of the network  $G = (V, E)$  (measured by  $|V|$  and  $|E|$ ), the width of the optical spectrum  $S$  (measured by  $|S|$ ), and the number of demands (measured by  $|K|$ ). Table I summarizes the order of the number of variables and constraints for the three models<sup>1</sup>.

TABLE I  
THE ORDER OF THE NUMBER OF VARIABLES AND CONSTRAINTS IN THE MODELS FROM THE LITERATURE.

	number of variables	number of constraints
model in [16]	$O( K ^2 E  S )$	$O( K ^2 E  S )$
model in [19]	$O( K ( E  +  S ))$	$O( K ^2 E  S )$
model in [1]	$O( K ( E  +  S  +  K ))$	$O( K ( E  +  V  +  K ))$

Our model uses also a polynomial number of variables, namely  $O(|K||E||S|)$ , but an exponential number of constraints due to

- path-continuity constraints (3) for all subsets  $X \subset V$  with  $o_k \in X, d_k \in V \setminus X$ , for all demands  $k \in K$ ,
- cycle-elimination constraints (5) for all subsets  $X_e \subset V$  containing both endnodes of edge  $e$ , for all edges  $e \in E$  and all demands  $k \in K$ .

Recall that path-continuity constraints (3) are used to force that the set  $E(k)$  of edges selected for the routing of demand  $k$  contains an  $(o_k, d_k)$ -path  $P_k$ , whereas cycle-elimination constraints (5) are used to prevent  $E(k)$  from containing cycles isolated from  $P_k$ , see Figure 2 for illustration. None of the

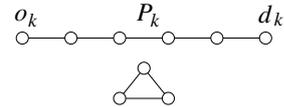


Fig. 2. A set  $E(k)$  containing an  $(o_k, d_k)$ -path  $P_k$  together with a cycle isolated from  $P_k$ .

models from [1], [16], [19] can exclude the occurrence of cycles isolated from  $P_k$ , the model presented in [19] can even not exclude cycles attached to  $P_k$ , see Figure 3 for illustration.

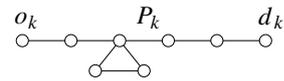


Fig. 3. A set  $E(k)$  containing an  $(o_k, d_k)$ -path  $P_k$  together with a cycle attached to  $P_k$ .

<sup>1</sup>To allow a comparison, we count integer variables with  $n$  possible values as  $n$  binary variables, and express variables encoding possible channels in terms of the spectrum width  $|S|$ .

In addition, none of the three models checks whether the transmission reach of routing paths is respected. Hence, all three models from [1], [16], [19] are incomplete as their feasible region is a superset of all feasible solutions of the RSA problem and can, thus, handle only some objective functions (where the optimal solution does neither contain cycles isolated from  $P_k$  nor cycles attached to  $P_k$ ).

Our model is the first complete edge-node formulation for the RSA problem as it precisely encodes the set of all feasible solutions, i.e., any integral vector satisfying all constraints from our model indeed corresponds to a feasible solution of the RSA problem. Therefore, our model can be used to optimize any objective function chosen as quality measure by the network operator.

In addition, our model is not only complete, but still tractable as we are able to separate the two exponentially-sized families of constraints (3) and (5) in polynomial time.

In fact by the polynomial equivalence between separation and optimization over rational polyhedra [5], the linear relaxations of our model can be solved in polynomial time if and only if the separation problem associated with inequalities (3) and (5) can be solved in polynomial time. The separation problem for the path-continuity constraints (3) reduces to  $O(|K|)$  minimum-cut problems in  $G$  and the separation problem for the cycle-elimination constraints (5) to  $O(|K||E|)$  minimum-cut problems in an auxiliary graph.

Therefore the separation problem associated with (3) and (5) is polynomially solvable using any polynomial-time maximum-flow algorithm (e.g., the preflow-push algorithm of Goldberg and Tarjan [4] running in  $O(|V|^3)$  time). Note that this separation approach provides the most-violated inequality if any w.r.t. a demand or a pair of a demand and an edge.

## V. COMPUTATIONAL RESULTS

In this section we present some preliminary computational results that mainly aim at assessing the empirical performances of a branch-and-cut framework based on our model for the three objectives functions presented in Section II and at comparing them with the results obtained by Velasco et al. [16] for  $O_2$  and by Cai et al. [1] for objective  $O_3$ .

In our experiments we therefore consider the Spanish Telefónica network represented in Figure 4 from [16] and three networks represented in Figure 5 from [1]. The characteristics of the topology of these four networks are given in Table II together with the available numbers of slots per link.

As none of the instances considered in [1], [16] were available, we randomly generated multisets of traffic demands, some of them using Net2Plan [12], while guaranteeing that some of those multisets share the properties described in [1], [16], that is, the same number of traffic demands (12 for Spanish Telefónica and 30 for n6s9) and the same range of values for the requested numbers of slots (in  $\{1, 2, 4\}$  for Spanish Telefónica and in  $\{1, \dots, 3\}, \dots, \{1, \dots, 9\}$  for n6s9). Table III summarizes the different types of traffic-demand multisets we considered for each network.

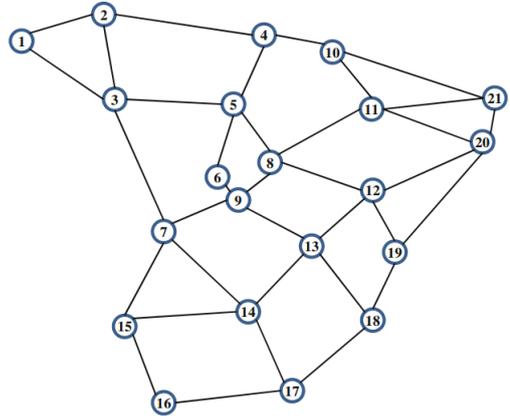


Fig. 4. Spanish Telefónica Network from [16]

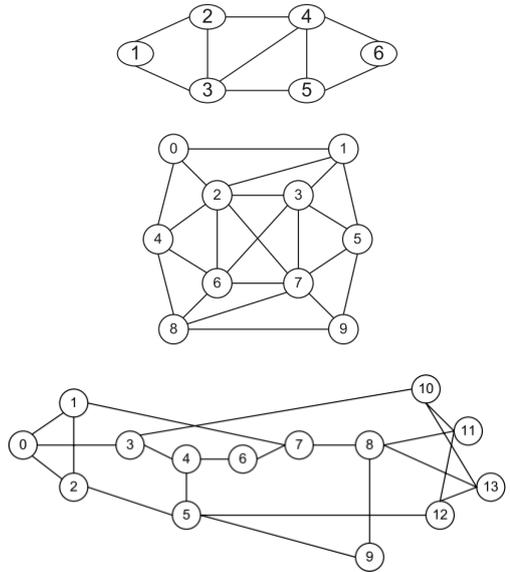


Fig. 5. n6s9, SmallNet, and NSFNET Networks from [1]

All our results were obtained on a laptop, running Microsoft Windows 10 Pro (64-bit), equipped with a 2.5GHz Intel Core i5-7300 HQ processor and 16-GB RAM. The branch-and-cut framework was implemented using IBM ILOG CPLEX Optimization Studio 12.8 C++ library. Note that using user-cut callbacks (needed for the separation of constraints (3) and (5)) in CPLEX 12.8 automatically deactivates the multithreading. To balance some struggles that the default heuristic of CPLEX has to generate good feasible solutions, we implemented a heuristic callback based on

- first decomposing for each demand  $k \in K$  its flow (given by the  $x_e^k$ -variables) into  $(o_k, d_k)$ -paths and
- second using a first-fit greedy approach to assign the best possible channels to the demands,

The first objective function  $O_1$  was considered in neither [1] nor [16]. Within a one-hour time limit, our branch-and-cut framework was able to solve to optimality all our instances

TABLE II  
CHARACTERISTICS OF THE NETWORK TOPOLOGIES

Network's name	number of nodes	number of links	number of slots per link
Spanish Telefónica	21	35	64
n6s9	6	9	80
SmallNet	10	22	{80, 100, 140, 180}
NSFNET	14	21	{120, 160, 210, 285}

TABLE III  
CHARACTERISTICS OF THE TRAFFIC DEMANDS

Network's name	number of demands	number of requested slots
Spanish Telefónica	{12, 15}	{1, 2, 4}
n6s9	{30, 50}	{1, ..., i}, i = 3, ..., 9
SmallNet	{100, 150, ..., 500}	{1, ..., 4}
NSFNET	{100, 150, ..., 250}	{2, ..., 6}

but the ones with 500 demands for which the optimality gap was under 0.5%. Over the course of the solution process, both the lower and upper bounds kept improving and only towards the end, optimal solutions were found.

For the second objective function  $O_2$ , Velasco et al. [16] were able to solve to optimality a single instance of Spanish Telefónica with 12 demands in over 6 hours. It took less than 3 hours for our branch-and-cut framework to solve to optimality the Spanish Telefónica instances with 12 demands and less than 6 hours for the Spanish Telefónica instances with 15 demands. We also ran our branch-and-cut framework on all the instances associated with n6s9 and were able to get optimal solutions within at most 15 minutes. Very early in the solution process, optimal solutions were found meaning that most of the solution time is dedicated to proving the optimality of those solutions (e.g., for the Spanish Telefónica with 12 demands, an optimal solution is found after about 15 minutes but proved optimal after about 2 hours and 40 minutes).

Cai et al. [1] only considered the third objective function  $O_3$  in their experiments with the additional property that given any two distinct nodes  $o$  and  $d$  of  $G$ , the multiset  $K$  of traffic demands contains either both demands having nodes  $o$  and  $d$  as their extremities (with the same requested number of slots) or none of them, and for the former case one assigned route is the reverse of the other one. Some of our generated instances for n6s9 fulfilled that property and were all solved to optimality within 20 minutes while Cai et al. [1] needed up to one hour to solve their similar instances (with CPLEX multithreading being active). We also ran our branch-and-cut framework on n6s9 instances without the reverse-demand property and for most of the instances were able to find optimal solutions within two hours and an optimality gap lower than 5% for the others. We noticed a similar behavior of the lower and upper bounds as for objective function  $O_1$ .

## VI. CONCLUDING REMARKS

The RSA problem in flexgrid elastic optical networks is an NP-hard problem for which various ILP models have been proposed in the literature. Hereby, detailed models aiming at precisely describing all technological aspects and being able to handle different criteria for optimization typically suffer from

tractability issues resulting from their greater complexity such that the tendency is to use simplified models.

The majority of the existing models uses edge-path formulations where the numbers of variables and constraints grow exponentially with the size of the instance, due to the huge number of feasible paths between all origin-destination pairs in the network. Hence, models based on edge-path formulations are often simplified by considering only subsets of precomputed paths (which cannot guarantee optimality, except for few objective functions) or require column-generation techniques (which limits the size of the instances that can be solved to optimality).

An alternative to edge-path formulations is to use edge-node formulations that have the advantage that the number of variables grows only polynomially with the size of the instance. Three compact edge-node formulations are presented in [1], [16], [19] where both the number of variables and constraints is polynomial in terms of the size of the instance. However, all three models are incomplete as their feasible region is a superset of all feasible solutions of the RSA problem and can, thus, handle only some objective functions.

Our contribution is to provide the first complete edge-node formulation for the RSA problem that precisely encodes the set of all feasible solutions and can, therefore, be used to optimize any chosen objective function. For that, we propose an appropriate combination of variables and constraints (partly using new variables and constraints) which results in a model having, as in [1], [16], [19], a polynomial number of variables, but an exponential number of constraints to ensure the exact encoding of feasible solutions.

As we are able to separate the exponentially-sized families of constraints in polynomial time, our model is computationally competitive with the compact but incomplete models from [1], [16], [19]. The computational results support this as our branch-and-cut solver was able, on the one hand, to efficiently handle larger instances and, on the other hand, to find optimal solutions for instances similar to those in [1], [16] in shorter time.

Hereby, we noticed by analyzing the computational results for objective function  $O_2$  that for most instances the optimal solution was found early in the computation process, but that most of the computation time was needed to certify its optimality. Hence, our future research also includes to strengthen lower bounds for the value of different objective functions in order to shorten the time during the computation needed for certifying optimality of a solution.

Therefore, we plan as future research, on the one hand, to strengthen our model further by devising new inequalities, e.g. derived as Chvátal-Gomory cuts from the initial constraints, and, on the other hand, to further improve the separation procedure for the exponentially-sized families of constraints.

Finally, recall that many different objective functions may be considered, depending on the network operator's choice. Besides  $O_1$ ,  $O_2$ ,  $O_3$ , the following objective functions may be of interest:

- $O_4$ : minimize the sum of the total length of paths (taking the edge weights  $l_e$  into account),
- $O_5$ : minimize the maximum load over all edges (where the load of an edge  $e$  is expressed by the number  $s_e$  of slots allocated on edge  $e$ ),
- $O_6$ : minimize the total cost of the solution (where the cost is expressed as the product of the length  $l_e$  and the load  $s_e$  of an edge  $e$ , summed up over all edges  $e$ ).

Hereby, the optimal solutions w.r.t. different objective functions may significantly differ such that an optimal solution for one objective may provide rather bad values according to other optimality criteria. For instance, the three optimal solutions presented in Example 2.1 ( $M_1$  for  $O_1$ ,  $M_2$  for  $O_2$ ,  $M_3$  for  $O_3$  which is also optimal for  $O_5$  minimizing the maximum edge load of 3) differ from each other and from the optimal solution for  $O_4$  and  $O_6$  presented in  $M_4$  (with minimum total length 13 of paths and minimum total cost 22).

$$M_4 = \begin{pmatrix} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ ab & 1 & 1 & & & 2 & & & & & \\ af & & & & & & & & & & \\ bc & 1 & 1 & 3 & 3 & 2 & 4 & & & & \\ cd & & & & & 2 & 4 & & & & \\ cf & & & 3 & 3 & & & & & & \\ de & 5 & 5 & 5 & & 4 & & & & & \\ df & & & & & & & & & & \\ ef & 5 & 5 & 5 & & & & & & & \end{pmatrix}$$

We notice that the objective functions

- $O_1, O_2, O_4$  for the routing may lead to solutions where some edges are highly loaded (with 6 slots in  $M_1, M_2$  and  $M_4$  where 3 slots suffice as in  $M_3$ ) which also forces a large used spectrum width (6 slots in  $M_1, M_2$  and  $M_4$  where 4 slots suffice as in  $M_3$ ),
- $O_3$  and  $O_5$  for the spectrum assignment may lead to routings along longer paths (total length of 17 in  $M_3$  where 13 suffice as in  $M_4$ ) which may also increase the total cost of the solution (29 for  $M_3$  where 22 suffice as in  $M_4$ ).

Hence, it is also in order to develop strategies to cope simultaneously with different quality measures of solutions.

#### ACKNOWLEDGMENT

This work was supported by the French National Research Agency grant ANR-17-CE25-0006, project FLEXOPTIM.

#### REFERENCES

- [1] Cai, A., G. Shen, L. Peng, M. Zukerman: *Novel Node-Arc Model and Multiiteration Heuristics for Static Routing and Spectrum Assignment in Elastic Optical Networks*, in: Journal of Lightwave Technology 2013, 3402-3413. DOI: 10.1109/JLT.2013.2282696
- [2] Christodouloupoulos, K., I. Tomkos and E. Varvarigos: *Elastic bandwidth allocation in flexible OFDM based optical networks*, IEEE J. Lightwave Technol. **29** (2011), 1354–1366. DOI: 10.1109/JLT.2011.2125777
- [3] Fayed, M., I. Katib, G.N. Rouskas and H.M. Faheem, *Spectrum Assignment in Mesh Elastic Optical Networks*, IEEE Proc. of ICCCN (2015), 1–6. DOI: 10.1109/ICCCN.2015.7288470
- [4] Goldberg, A.V., and Tarjan, R.E.: *A New Approach to the Maximum Flow Problem*, in: Journal on the Association for Computing Machinery. **35** (1988), 921–940.
- [5] Gröschel, M., Lovász, L., and Schrijver, A.: *Geometric Algorithms and Combinatorial Optimization*, Springer-Verlag, 1988.
- [6] Jia, X., F. Ning, S. Yin, D. Wang, J. Zhang and S. Huang: *An integrated ILP model for Routing, Modulation Level and Spectrum Allocation in the next generation DCN*, in: Third International Conference on Cyberspace Technology (CCT) 2015, 1–3. DOI: 10.1049/cp.2015.0821
- [7] M. Jinno, H. Takara, B. Kozicki, Y. Tsukishima, Y. Sone, and S. Matsuoka: *Spectrum-efficient and scalable elastic optical path network: architecture, benefits, and enabling technologies*, IEEE Commun. Mag. **47** (2009) 66–73. DOI: 10.1109/MCOM.2009.5307468
- [8] Klinkowski, M., J. Pedro, D. Careglio, M. Pióro, J. Pires, P. Monteiro, and J. Solé-Pareta: *An overview of routing methods in optical burst switching networks*, in: Optical Switching and Networking 2010, 41 – 53. DOI: doi.org/10.1016/j.osn.2010.01.001
- [9] Klinkowski, M., and K. Walkowiak: *Routing and Spectrum Assignment in Spectrum Sliced Elastic Optical Path Network*, in: IEEE Communications Society 2011, 884 - 886. DOI: 10.1109/LCOMM.2011.060811.110281
- [10] Klinkowski, M., M. Pióro, M. Zotkiewicz, K. Walkowiak, M. Ruiz, and L. Velasco: *Spectrum allocation problem in elastic optical networks - a branch-and-price approach*, in: 17th International Conference on Transparent Optical Networks (ICTON) 2015, 1–5. DOI: 10.1109/ICTON.2015.7193482
- [11] Klinkowski, M. and K. Walkowiak: *A Simulated Annealing Heuristic for a Branch and Price-Based Routing and Spectrum Allocation Algorithm in Elastic Optical Networks*, in: Intelligent Data Engineering and Automated Learning Book 2015, 290–299. DOI: 10.1007/s11107-012-0378-7
- [12] Net2Plan [www.net2plan.com/ocn-book](http://www.net2plan.com/ocn-book).
- [13] Ruiz, M., and M. Pióro, M. Zotkiewicz, and M. Klinkowski and L. Velasco: *Column generation algorithm for RSA problems in flexgrid optical networks*, in: Photonic Network Communications 2013, 53–64. DOI: 10.1007/s11107-013-0408-0
- [14] Shirazipourazad, S., Ch. Zhou, Z. Derakhshandeh and A. Sen: *On routing and spectrum allocation in spectrum sliced optical networks*, Proceedings of IEEE INFOCOM (2013), 385–389. DOI: 10.1109/INFCOM.2013.6566800
- [15] Talebi, S., F. Alam, I. Katib, M. Khamis, R. Salama, and G.N. Rouskas: *Spectrum management techniques for elastic optical networks: A survey*, Optical Switching and Networking **13** (2014), 34–48. DOI: doi.org/10.1016/j.osn.2014.02.003
- [16] Velasco, L., M. Klinkowski, M. Ruiz, and J. Comellas: *Modeling the routing and spectrum allocation problem for flexgrid optical networks*, in: Photonic Network Communications 2012, 177–186. DOI: 10.1007/s11107-012-0378-7
- [17] Walkowiak, K., R. Gosien, M. Klinkowski, and M. Wozniak: *Optimization of Multicast Traffic in Elastic Optical Networks With Distance-Adaptive Transmission*, in: IEEE Communications Letters 2014, 2117–2120. DOI: 10.1109/LCOMM.2014.2367511
- [18] Wang, Y., X. Cao and Y. Pan: *A study of the routing and spectrum allocation in spectrum-sliced elastic optical path networks*, in: Proceedings of IEEE INFOCOM 2011, 1503-1511. DOI: 10.1109/INFCOM.2011.5934939
- [19] Zotkiewicz, M., M. Pióro, M. Ruiz, M. Klinkowski, and L. Velasco: *Optimization models for flexgrid elastic optical networks*, in: 15th International Conference on Transparent Optical Networks (ICTON) 2013, 1–4. DOI: 10.1109/ICTON.2013.6602691

# An Efficient Exhaustive Search for the Discretizable Distance Geometry Problem with Interval Data

A. Mucherino\*, J-H. Lin†

†IRISA, University of Rennes 1, Rennes, France.  
antonio.mucherino@irisa.fr

‡Research Center for Applied Sciences, Academia Sinica, Taipei, Taiwan.  
jhlin@gate.sinica.edu.tw

**Abstract**—The Distance Geometry Problem (DGP) asks whether a simple weighted undirected graph can be realized in a given space (generally Euclidean) so that a given set of distance constraints (associated to the edges of the graph) is satisfied. The Discretizable DGP (DDGP) represents a subclass of instances where the search space can be reduced to a discrete domain having the structure of a tree. In the ideal case where all distances are precise, the tree is binary and one singleton, representing one possible position for a vertex of the graph, is associated to every tree node. When the distance information is however not precise, the uncertainty on the distance values implies that a three-dimensional region of the search space needs to be assigned to some nodes of the tree.

By using a recently proposed coarse-grained representation for DDGP solutions, we extend in this work the branch-and-prune (BP) algorithm so that it can efficiently perform an exhaustive search of the search domain, even when the uncertainty on the distances is important. Instead of associating singletons to nodes, we consider a pair consisting of a box and of a most-likely position for the vertex in this box. Initial estimations of the vertex positions in every box can be subsequently refined by using local optimization.

The aim of this paper is two-fold: (i) we propose a new simple method for the computation of the three-dimensional boxes to be associated to the nodes of the search tree; (ii) we introduce the resolution parameter  $\rho$ , with the aim of controlling the similarity between pairs of solutions in the solution set. Some initial computational experiments show that our algorithm extension, differently from previously proposed variants of the BP algorithm, is actually able to terminate the enumeration of the solution set by providing solutions that differ from one another accordingly to the given resolution parameter.

## I. INTRODUCTION

LET  $G = (V, E, d)$  be a simple weighted undirected graph, where vertices represent objects (whose nature depends on the application at hand), and the existence of an edge  $\{u, v\}$  between two vertices  $u$  and  $v$  indicates that the distance between the two corresponding objects is known [18]. The weight  $d(u, v)$  associated to the edge  $\{u, v\}$  is in general a real-valued interval providing the lower and the upper bound on the distance values. However, the interval  $d(u, v)$  can degenerate to one singleton, and in this situation only one approximation of the distance value is available.

**Definition 1** Given a simple weighted undirected graph  $G = (V, E, d)$  and a positive integer  $K$ , the Distance Geometry Problem (DGP) asks whether a function

$$x : v \in V \longrightarrow x_v \in \mathbb{R}^K$$

exists such that

$$\forall \{u, v\} \in E, \quad \|x_u - x_v\| \in d(u, v), \quad (1)$$

where  $\|\cdot\|$  represents the Euclidean norm.

The function  $x$  is called a *realization* of the graph  $G$ . We say that a realization  $x$  that satisfies all constraints in equ. (1) is a *valid realization*.

The DGP is NP-hard [26], and has several different applications, including: (i) protein structure determination [7] (this is the application we will consider in our experiments in Section IV); (ii) sensor network localization [27]; (iii) multi-dimensional scaling [13]; (iv) clock synchronization [8]; (v) motion adaptation [25]; and others.

We give the following definition of a discretizable subclass of DGP instances [23]. Let  $E'$  be the subset of the edge set  $E$  such that the weight associated to the edges are degenerate intervals.

**Definition 2** A simple weighted undirected graph  $G$  represents an instance of the Discretizable DGP (DDGP) if and only if there exists a vertex ordering on  $V$  such that the following two assumptions are satisfied:

- (a)  $G[\{1, 2, \dots, K\}]$  is a clique whose edges are in  $E'$ ;
- (b)  $\forall v \in \{K + 1, \dots, |V|\}$ , there exist  $u_1, u_2, \dots, u_K \in V$  such that
  - (b.1)  $u_1 < v, u_2 < v, \dots, u_K < v$ ;
  - (b.2)  $\{\{u_1, v\}, \{u_2, v\}, \dots, \{u_{K-1}, v\}\} \subset E'$ ,  
 $\{u_K, v\} \in E$ ;
  - (b.3)  $\mathcal{V}_S(u_1, u_2, \dots, u_K) > 0$  (if  $K > 1$ ),

where  $G[\cdot]$  is the subgraph induced by a subset of vertices of  $V$ , and  $\mathcal{V}_S(\cdot)$  is the volume of the simplex generated by a valid realization of the vertices  $u_1, u_2, \dots, u_K$ .

In the following, we will refer to assumptions **(a)** and **(b)** as the *discretization assumptions*. Such assumptions can be verified only if a vertex ordering is associated to  $V$  [10], which is generally referred to as a *discretization order* when the two assumptions above are satisfied.

Assumption **(a)** allows us to fix a coordinate space where to construct DDGP solutions while making sure that none of them can be obtained from other solutions by applying translations or rotations. Assumption **(b)** ensures that every vertex  $v$  has at least  $K$  *reference vertices*  $u_i$ , with  $1 \leq i \leq K$ , such that the corresponding *reference distance* to  $v$  is known. Since it is supposed (see Assumption **(b.2)**) that only one distance is represented by a non-degenerate interval, the possible positions for  $v$  wrt its reference vertices can be obtained by intersecting  $K - 1$  spheres and one spherical shell, which gives at most two arcs [15]. These assumptions make it possible to reduce the DDGP search space to a discrete domain having the structure of a tree where (possibly degenerate) arcs are associated to its nodes.

In order to simplify the notations, and in accordance with the application that is considered in Section IV, we will suppose in the following that the dimension  $K$  is set to 3.

The branch-and-prune (BP) algorithm [17] can be employed for exploring the search tree obtained with the discretization. In a recent work, we have integrated the BP algorithm with a coarse-grained representation [24]. This representation allows us to deal in an efficient way with the uncertainty of the available distance values, which can have an important impact on the lengths of the arcs obtained with the intersections. Differently from [15], where sample points are selected from the arcs, the coarse-grained representation better deals with uncertainty by assigning to every node of the search tree not only a suitable position  $x_v$  for the corresponding vertex  $v$ , but also a three-dimensional region  $B_v$  where  $v$  is allowed to take its positions. While the initial estimation for  $x_v$  can be very rough, the region  $B_v$  contains all its feasible positions and can therefore be explored for *refining* the position  $x_v$  while searching in a relatively small neighborhood of the search domain. In our first studies, this three-dimensional region is represented by a box inscribing the arcs obtained with the intersections.

This work is a step ahead in the development of an implementation of the BP algorithm that is based on the coarse-grained representation. Our new implementation is the first one that is actually able to enumerate the entire solution set of DDGP instances containing approximated distances (see Section IV). To this final purpose, we propose the integration in the algorithm of the following two features:

- a simple strategy for the definition of the boxes inscribing the arcs obtained with the intersections of the spheres (degenerate intervals) and spherical shell (one non-degenerate interval);
- the introduction of the resolution parameter, which allows to neglect “on-the-fly” all solutions that are *too similar* to solutions that were already computed.

The rest of the paper is organized in three main sections. Section II will be focused on our implementation of the BP algorithm: we will describe the coarse-grained representation, as well as our new method to compute the boxes inscribing the arcs obtained with the discretization process. Section III will introduce the resolution parameter and discuss its impact on the execution of the BP algorithm. Finally, Section IV will present some experiments on DDGP instances of the protein structure determination problem, while Section V will conclude the paper.

## II. AN EXTENDED BP ALGORITHM

We have recently proposed the use of a coarse-grained representation of the DDGP search space in [24]. In the present work, we will extend this approach by introducing some new features in the BP algorithm, so that a *complete* enumeration of the search space will in fact be possible, even in presence of interval distances. This was the main objective of various previous publications (see for example [6]), but it was not completely attained.

In the discussion below, we will focus on the following main points. A general sketch of the BP algorithmic framework will be given in Section II-A, while the coarse-grained representation will be detailed in Section II-B. Then, Section II-C will discuss on how arcs of vertex positions can be computed by exploiting the available distance information, and Section II-D will present our method for the definition of boxes inscribing the arcs.

### A. The BP algorithm

The BP algorithm was formally introduced in [17], and its basic idea is to perform a systematic exploration of the DDGP search tree. This search domain can be explored starting from its top, where the first vertex belonging to the initial clique is placed. Subsequently, all other vertices in the initial clique can be placed in their unique positions [14], and then the search can actually start with the vertex having rank 4 in the associated discretization order. At each step, the candidate positions for the current vertex  $v$  are computed, and the search is branched. This phase of the BP algorithm is named *branching phase*. Depending on the available distance information, represented by one approximated value, or rather by a real-valued interval, the set of candidate positions may either contain two singletons, or two disjoint arcs, respectively. Therefore, an arc is in general associated to every tree node, which can be in some cases degenerate. The distances that are used during the branching phase are called “discretization distances”.

Pruning devices can be employed for discovering infeasible vertex positions. In BP, the main pruning device verifies whether available distances, that are not used for the discretization, are satisfied by candidate vertex positions or not. As soon as a vertex position is found to be infeasible, then the corresponding branch can be pruned and the search can be backtracked [16]. This phase of the BP algorithm is named

**Algorithm 1** The BP algorithm's main framework

---

```

1: BP( $v, G$ )
2: if ( $v > |V|$ ) then
3:   // one solution is found
4:   print current conformation;
5: else
6:   // coordinate computation
7:   if (one discretization distance belongs to  $E \setminus E'$ ) then
8:     compute two candidate arcs
9:     add them to the list  $L$ 
10:  else
11:    compute two candidate positions  $y^1$  and  $y^2$ 
12:    add them to the list  $L$ 
13:  end if
14:  for  $i = 1, \dots, |L|$  do
15:    if ( $L(i)$  is an arc) then
16:      choose sample  $x_v$  from the arc  $L(i)$ 
17:    else
18:      set  $x_v = y^i$ 
19:    end if
20:    // verifying the feasibility of the computed positions
21:    if ( $x_v$  is feasible) then
22:      BP( $v + 1, G$ );
23:    end if
24:  end for
25: end if

```

---

*pruning phase*, and the used distances are called “pruning distances”.

Algorithm 1 is a sketch of the main framework for the BP algorithm. In the BP call,  $v \in V$  is the current vertex to be positioned and  $G$  is the simple weighted undirected graph representing a DDGP instance. Once the initial clique has been realized, the BP algorithm can be invoked recursively, starting from the vertex  $v$  having rank 4. As mentioned above, a lot of research has been conducted in recent years to find the best way to implement the line 16 of the algorithm. In [15], for example, a predefined number  $D$  of sample points are extracted from the generated arcs (the parameter  $D$  is the “discretization factor”). However, this strategy for a *lossy discretization* of the arcs has an important impact on the quality of the solutions, with a consequent amplification of error propagation along the search tree [12].

In our current implementation of the BP algorithm, we do not discretize the arcs, but we rather consider the coarse-grained representation presented in Section II-B. Only one vertex position is associated to every node of the tree, but this position is not fixed in one unique position. If necessary, it can rather be refined subsequently when deeper layers of the tree are reached, by exploring other possible positions inside the box that is associated to the node. It is important to remark that, when the generated arcs are larger, the corresponding box becomes bigger, and it might include positions that are feasible with more than one solution.

*B. A coarse-grained representation for BP*

Previous attempts to improve the efficiency of the BP algorithm (see for example [1], [11]) were based on the idea to avoid branching over subsets of positions from the arcs that may be found to be infeasible at the current layer *before* starting the branching phase. While some improvements were observed, these BP variants are however not able to consider distances that appear subsequently at further tree layers.

This is the main motivation for a coarse-grained representation of DDGP solutions. Instead of fixing, on every branch of the tree, all vertices in unique positions, the idea is to rather associate a small *region* of the search space to every vertex, together with a most-likely position. The shape of the region can be chosen on the basis of the methods that are then used for their manipulation.

In our coarse-grained representation, we use the following function:

$$z : v \in V \longrightarrow (x_v, B_v) \in \mathbb{R}^3 \times \mathbb{R}^6,$$

where  $B_v$  is a box defined in the Cartesian system given by the initial clique (see Section II-A). We point out that  $B_v$  has 6 dimensions (in dimension  $K = 3$ , the position of one vertex of the box, plus the corresponding depth, length and height values are necessary for its unique determination). When a new vertex position  $x_v$  is generated for the current vertex  $v$ , the function  $z$  does not only allow to assign a position  $x_v$  to  $v$ , but also to keep track of the feasible region where it belongs to. On further layers, in fact, the position  $x_v$  may not be feasible wrt some other distances, and it could therefore be *refined* in order to ensure global feasibility. This can be done, for example, by employing solvers for local optimization. The position  $x_v$  is naturally constrained to stay in the original box  $B_v$  for two main reasons. Firstly, the (continuous) search space of the local solver is in this way bounded; secondly, the situation where the local solver can move to solutions belonging to other tree branches is avoided.

We motivate the choice of employing a local solver with the fact that, at every layer of the tree where an infeasibility is discovered, there are only a few distances that are not satisfied, and the actual search space consists of the set product of all boxes  $B_v$ . This makes the corresponding subproblem to solve easier to tackle. Naturally, an important point concerning the use of a local solver is also its fast convergence: in fact, when attempting the solution of harder instances, we expect the local solver to be invoked at almost all recursive BP calls.

In this work, we will use a Spectral Projected Gradient (SPG) method [3], [22], [24] for this refinement step. When the BP algorithm reaches a leaf node, a valid realization  $x$  can be extracted from  $z$  by simply extracting the set of positions  $x_v$ , for every  $v \in V$ .

*C. Computing arcs of vertex positions*

When the discretization assumptions are satisfied, the possible positions for a given vertex  $v$  can be computed by exploiting the set of discretization distances, together with the positions, along the same tree branch, of the corresponding

reference vertices. Let  $u_3$ ,  $u_2$ , and  $u_1$  be the three reference vertices for the vertex  $v \in V$ . Since all reference vertices precede  $v$  in the given discretization order, one position for every  $u_i$  is assigned to the current tree branch, and distances between pairs of reference vertices  $u_i$  may be computed (if not already available). Therefore, the subgraph  $G[\{u_3, u_2, u_1, v\}]$ , when completed with missing distances between reference vertices, always induces a clique. As shown in [14], the distance information in the clique can be exploited to represent the possible positions for  $v$  in terms of torsion angles  $\omega$ . More precisely, once the distances  $d(u_3, u_2)$ ,  $d(u_2, u_1)$ ,  $d(u_1, v)$  are fixed, as well as the angles  $\theta_{u_2}$  and  $\theta_{u_1}$  formed, respectively, by the triplets  $(u_3, u_2, u_1)$  and  $(u_2, u_1, v)$ , then the distance  $d(u_3, v)$  corresponds to two possible values for the torsion angle  $\omega$  formed by the quadruplet  $(u_3, u_2, u_1, v)$  [19].

Let us initially suppose that the distance  $d(u_3, v)$  is exact: at most two distinct values  $\omega^+$  and  $\omega^-$  can be computed for the torsion angle (it might happen that they coincide). The two corresponding positions for the vertex  $v$  can therefore be computed with the following formula:

$$\chi(v, \omega) = x(u_1) + Uw(v, \omega), \quad (2)$$

where the matrix  $U$  is a rotation matrix (see [9]) and

$$w(v, \omega) = \begin{bmatrix} -d(u_1, v) \cos \theta_{u_1}, \\ d(u_1, v) \sin \theta_{u_1} \cos \omega, \\ d(u_1, v) \sin \theta_{u_1} \sin \omega. \end{bmatrix}. \quad (3)$$

When the distance  $d(u_3, v)$  is represented by an interval, two intervals of  $\omega$  can be defined [11]. From a geometrical point of view, the interval distance  $d(u_3, v)$  corresponds to two arcs that can be identified over the circle obtained by intersecting two spheres centered in  $u_1$  and  $u_2$  and having as radius, respectively, the corresponding reference distances. We generate a three-dimensional box inscribing every obtained arc, so that it can then be associated to the tree node, together with one position taken from the arc. This chosen position is the one that will be used at further layers when defining spheres or spherical shells for the intersections; however, this position may need to be refined when infeasibilities are detected by the pruning devices. The refinement step is supposed to keep the given position  $x_v$  inside the predetermined box  $B_v$ .

#### D. Computing the boxes

The boxes are defined by using the minimal and maximal possible coordinates associated to a given interval for  $\omega$ . To perform this calculation, we remark that the three components of  $\chi(v, \omega)$  (see equ. (2) and equ. (3)) can be rewritten (we explicitly write only the first component  $\chi'$ ) as:

$$\chi'(v, \omega) = A(v) \cos \omega + B(v) \sin \omega + C(v),$$

where  $C(v)$  is an additive term, and  $A(v)$  and  $B(v)$  are two multiplicative terms. For fixed  $v \in V$ , since  $C(v)$  is an additive constant, the determination of the minimal and maximal values for the first component of  $\chi(v, \omega)$  can be focused on the

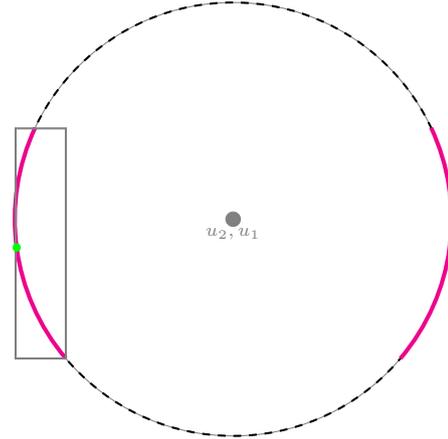


Fig. 1. In dashed line, the circle obtained by intersecting the two spheres centered in  $u_1$  and  $u_2$ . In purple, the arcs on the circle obtained by intersecting this circle with the spherical shell centered in  $u_3$ . For one arc, we show the box inscribing the arc in gray, together with the arc central position in green.

optimization of the remaining terms. By following [11], we remark that

$$A(v) \cos \omega + B(v) \sin \omega = R(v) \cos(\omega - \alpha),$$

where the pair  $(R(v), \alpha)$  corresponds to the polar coordinates of  $(A(v), B(v))$  in dimension 2:

$$\begin{cases} A(v) = R(v) \cos \alpha, \\ B(v) = R(v) \sin \alpha. \end{cases}$$

As a consequence, the problem of finding the minimal and maximal values for  $\chi'(v, \omega)$  reduces to the one of finding the optimal values for the cosine function. Notice that the value of  $\alpha$  can be computed as

$$\alpha = \arctan\left(\frac{B(v)}{A(v)}\right)$$

and that the same strategy can be used for finding the minimal and maximal values for the other components of  $\chi(v, \omega)$  over the given interval for  $\omega$ . A graphical representation of the obtained boxes is given in Fig. 1.

### III. RESOLUTION PARAMETER

At every recursive call of the BP algorithm (see Section II-A), the intersection of two spheres with one spherical shell produces two arcs [11]. The coarse-grained representation (see Section II-B) replaces every arc with a box inscribing the arc (see Section II-D) and a most-likely position, which is initially set at the arc central point. Every generated pair  $(x_v, B_v)$  consisting of a position and a box for the vertex  $v$  can be then assigned to the nodes of the search tree.

The *resolution* parameter  $\rho$  is integrated in the BP algorithm for controlling the size of the boxes associated to the tree nodes. This is done at two different levels:

- 1) If the length of the current arc is larger than the resolution parameter  $\rho$ , then the arc is split in a sufficient number of equally-long sub-arcs whose length is smaller

than  $\rho$ . Naturally, this strategy implies that the search domain is not binary anymore at all layers, for many vertices may need more than two arcs for satisfying the required resolution. On the one hand, this procedure increases the complexity of the search; on the other hand, it allows to assign smaller boxes to the tree nodes, so that the search domain of the local optimizer becomes smaller as well.

- 2) If at least one solution has already been found, and the BP algorithm is currently exploring alternative nodes at a given layer  $v$ , then only positions  $x_v$  whose Euclidean distance from the position of  $v$  in the previously found solution is larger than  $\rho$  are considered. In fact, when this distance is smaller than  $\rho$ , the previous and the current solution are “too” similar, and the current one can therefore be discarded. Notice that, even if this Euclidean distance may be larger than  $\rho$  when the nodes are generated, the local optimizer may modify the positions  $x_v$  so that it subsequently becomes smaller than  $\rho$ . In such a case, the previous and the current solution have the tendency to converge to the same conformation, and thus the current branch can be discarded.

To sum up, the resolution parameter  $\rho$  does not only influence the branching phase of the BP algorithm, but it rather performs two kinds of verification during the execution of the algorithm ensuring that all generated solutions differ from one another in accordance with the chosen resolution parameter.

#### IV. COMPUTATIONAL EXPERIMENTS

We present in this section some initial computational experiments on a set of artificially generated instances. All codes were written in C programming language and all experiments were carried out on an Intel Core i7 2.30GHz with 8GB RAM, running Linux. The codes have been compiled by the GNU C compiler v.4.9.2 with the `-O3` flag.

Before showing our computational experiments, we will briefly present the considered application in structural biology concerning the determination of protein structures (see Section IV-A), and we will explain how we generated our instances (see Section IV-B). Section IV-C will present the experiments.

##### A. Protein structure determination

One of the classical applications of the DGP arises in structural biology [7]. Distances between atom pairs in a given molecule can be estimated through experiments of Nuclear Magnetic Resonance (NMR), so that the possible conformations of the molecule in the three-dimensional space can be identified by solving an instance of the DGP. This application is of relevant interest, especially when dealing with proteins, because the identification of protein conformations can give insights on the dynamics of such molecules, and therefore on their function.

It was proved that protein instances of the DGP belong to the subclass of the DDGP [14], [23]. In many papers cited above (see for example [15], [6], [12]), protein instances are

TABLE I  
SOME EXPERIMENTS ON PROTEIN INSTANCES RESEMBLING NMR DATA.

<i>protein</i>	$ V $	$ E $	$ E' $	$\rho$	#sols	best MDE	time
2jmy	77	428	219	0.5	6	1.73e-05	1m 38s
				1.0	3	1.90e-05	54s
				2.0	2	2.40e-05	51s
2kxa	121	700	367	2.0	2	3.14e-05	45m 28s
				3.0	1	9.94e-05	7m 31s
2ksl	254	1388	684	2.0	2	2.42e-05	16m 55s
				2.0	1	3.47e-05	4m 5s

used to perform the experiments. However, as already pointed out in the Introduction, none of such previous works were able to perform an exhaustive search on the domain of the considered instances. Our experiments will show that the BP algorithm, integrated with the new features introduced in this work, is actually able to perform this exhaustive search.

##### B. Generation of the instances

We selected the protein conformations that were considered in the experiments presented in [6] and [24]. We do not use real NMR data, but we rather generate our protein instances from known models of the selected proteins. The three considered proteins, having codes 2jmy, 2kxa and 2ksl in the Protein Data Bank (PDB) [2], have been experimentally determined by NMR experiments, and, as it is usually the case, more than one model for each protein was deposited. In our instance generation, we have simply considered the first model that appears in the corresponding PDB file.

Our instances are generated in a way to resemble NMR data. From the initial conformation model, we compute all inter-atomic distances, and we include in our instance the following distances:

- distances between bonded atoms (only one real value approximated to 3 decimal digits);
- distances between atoms bonded to a common atom (only one real value approximated to 3 decimal digits);
- distances between the first and the last atom forming a torsion angle (distances represented by an interval);
- distances between hydrogen atoms that are shorter than 5Å (distances represented by an interval).

In order to define the interval distances, we create an interval of range  $0.1\text{\AA}$  for the distances related to torsion angles, and an interval of range  $0.5\text{\AA}$  for distances related to hydrogens, and we place the true distance randomly inside such an interval. The atoms are sorted accordingly to the order proposed in [21], which ensures the discretizability of the instance.

##### C. Some initial experiments

We present some initial experiments performed by considering the instances generated as detailed in the previous section. Table I summarizes our experiments: the information about the graph representing the DDGP instance is given together with the chosen resolution  $\rho$ . Moreover, for every experiment, the

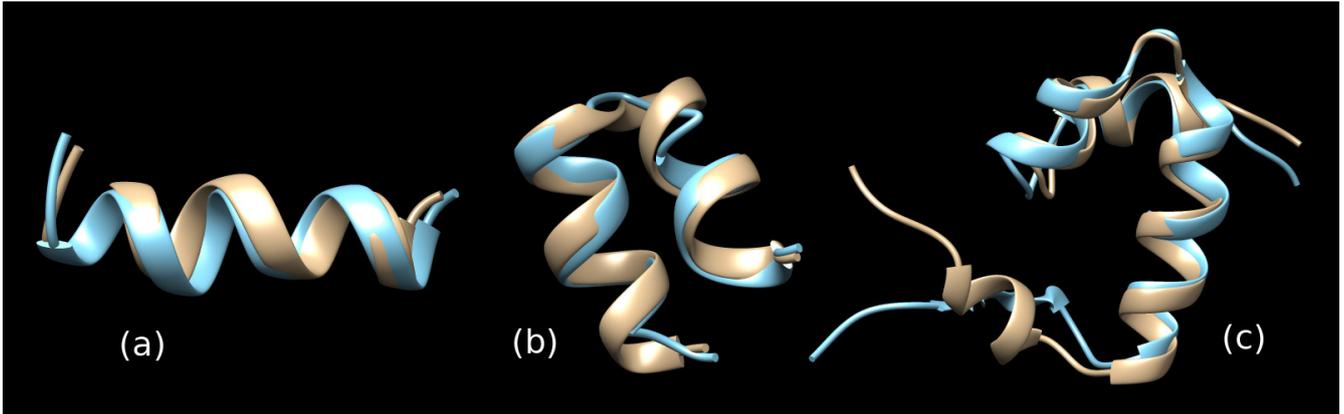


Fig. 2. For every considered protein instance (see Table I), we propose the alignment between the protein model (in gold) that we considered for the generation of the instances, and the best matching solution found by our BP algorithm implementation (in blue). (a) is the alignment obtained for 2jmy; (b) for 2kxa; (c) for 2ks1. The BP solution was subject to energy minimization before the alignment.

total number (#sols) of found solutions is reported, together with the best MDE value:

$$\text{MDE}(x) = \frac{1}{|E|} \sum_{\{u,v\} \in E} \frac{\Delta(\|x_u - x_v\|, d(u,v))}{d(u,v)}, \quad (4)$$

and with the time in minutes (m) and seconds (s). We point out that the given number #sols of solutions is the value of our solution counter at the end of the *complete* enumeration of the search tree, with the specified resolution parameter  $\rho$ . The function  $\Delta(y, I)$  in equ. (4) computes the distance on a line between the real value  $y$  and the real-valued interval  $I$ .

SPG is invoked as a local optimization solver for performing the refinement step (see Section II-B), with the same general settings used in [22]. SPG can terminate because of different criteria: either when the objective function value becomes smaller than  $10^{-6}$ , or when the norm of the search direction becomes smaller than  $10^{-6}$ , or when it reaches the maximum number of allowed interactions, which is set to 10000 in our experiments. Notice that the objective function optimized by SPG is not the MDE function above (MDE is in fact not differentiable in its entire definition domain): more information about SPG and its implementation can be found in [22], [25].

It is easy to see that the newly introduced resolution parameter is able to control the cardinality of the final solution set. The more its value is large, the less are the found solutions (i.e. more solutions are discarded because considered to be too similar to previously obtained solutions). The resolution parameter also influences the total computational time, because it allows to skip all branches potentially leading to similar solutions.

In order to verify how the BP algorithm is able to reconstruct the original protein models used to generate our instances, we have aligned the original structure with the obtained solutions. Before alignment, however, for the two compared structures to be in the same conditions, we optimized the internal energy of the BP structures. To perform such an optimization, the topology/parameter file and the coordinate

file were prepared by the `tLEaP` module of the AMBER 16 program suite [5]. The `GBn` model of Mongan et al. [20] was used for the implicit solvent model; the Bondi radii set was also used [4]. 250 steps of steepest descent minimization were followed by 250 steps of conjugated gradient minimization. The MPI version of `pmemd` program of AMBER 16 suite was used for energy minimization.

Fig. 2 shows the obtained alignments. They show that the BP algorithm is actually able to *reconstruct* the original protein model that was used to generate our instances. We can remark moreover that the quality of the solutions, measured through the MDE function, is independent on the resolution parameter, and has a rather constant magnitude in all experiments. Its value indicates a very good quality for the found solutions (recall that the distances represented by only one approximated value are represented with 3 decimal digits).

## V. CONCLUSION

We have presented an extended version of the BP algorithm which allows an efficient exploration of the search tree obtained with the discretization of the DGP. This extended version is in fact capable of enumerating exhaustively the search tree even when the distance information is given through real-valued intervals. A pair consisting of a three-dimensional box and a selected vertex position in the box is associated to every node of the tree, so that the selected position can be refined at further layers of the search tree when new distance information needs to be verified. The inclusion of a resolution parameter allows to generate boxes with controlled sizes, and to perform the BP branching phase only when the new added branches lead to the generation of solutions that differ, in accordance with the resolution parameter, from solutions that were already computed.

One of the first objectives of our future works will consist in solving DGP instances containing NMR data that are obtained through the experimental technique. To this aim, the main challenge that we will need to face is given by the lower

precision of the distance information. In fact, the intervals related to distances derived from NMR experiments may correspond to ranges up to 3Å. Moreover, the possibility to skip the energy minimization step (that was performed in our computational experiments) will be studied by including more stringent distance constraints for important hydrogen bonding distances.

Finally, work will be performed for formalizing the concepts related to the introduction of the resolution parameter, which will require a complete understanding of the actual impact of this new parameter on the BP algorithm.

#### ACKNOWLEDGMENTS

Most of this work was performed during the visit of AM to Academia Sinica (April 2019), and of JHL to IRISA (May 2019), which were made possible thanks to our CNRS-MoST PRC project entitled “Rapid NMR Protein Structure Determination and Conformational Transition Sampling by a Novel Geometrical Approach” (years 2018–19).

#### REFERENCES

- [1] R. Alves, C. Lavor, *Geometric Algebra to Model Uncertainties in the Discretizable Molecular Distance Geometry Problem*, *Advances in Applied Clifford Algebra* **27**, 439–452, 2017.
- [2] H. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. Bhat, H. Weissig, I. Shindyalov, P. Bourne, *The Protein Data Bank*, *Nucleic Acids Research* **28**, 235–242, 2000.
- [3] E.G. Birgin, J.M. Martínez, M. Raydan, *Spectral Projected Gradient methods: Review and Perspectives*, *Journal of Statistical Software* **60**(i03), 21 pages, 2014.
- [4] A. Bondi, *van der Waals Volumes and Radii*, *Journal of Physical Chemistry* **68**(3), 441–451, 1964.
- [5] D.A. Case, R.M. Betz, D.S. Cerutti, T.E. Cheatham III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, H.T. Nguyen, I. Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, L. Xiao, P.A. Kollman, *AMBER 2016*, University of California, San Francisco, 2016.
- [6] A. Cassioli, B. Bardiaux, G. Bouvier, A. Mucherino, R. Alves, L. Liberti, M. Nilges, C. Lavor, T.E. Malliavin, *An Algorithm to Enumerate all Possible Protein Conformations verifying a Set of Distance Restraints*, *BMC Bioinformatics* **16**:23, 15 pages, 2015.
- [7] G.M. Crippen, T.F. Havel, *Distance Geometry and Molecular Conformation*, John Wiley & Sons, 1988.
- [8] N.M. Freris, S.R. Graham, P.R. Kumar, *Fundamental Limits on Synchronizing Clocks Over Networks*, *IEEE Transactions on Automatic Control* **56**(6), 1352–1364, 2010.
- [9] D.S. Gonçalves, A. Mucherino, *Discretization Orders and Efficient Computation of Cartesian Coordinates for Distance Geometry*, *Optimization Letters* **8**(7), 2111–2125, 2014.
- [10] D.S. Gonçalves, A. Mucherino, *Optimal Partial Discretization Orders for Discretizable Distance Geometry*, *International Transactions in Operational Research* **23**(5), 947–967, 2016.
- [11] D.S. Gonçalves, A. Mucherino, C. Lavor, *An Adaptive Branching Scheme for the Branch & Prune Algorithm applied to Distance Geometry*, *IEEE Conference Proceedings, Federated Conference on Computer Science and Information Systems (FedCSIS14), Workshop on Computational Optimization (WCO14)*, Warsaw, Poland, 463–469, 2014.
- [12] D.S. Gonçalves, A. Mucherino, C. Lavor, L. Liberti, *Recent Advances on the Interval Distance Geometry Problem*, *Journal of Global Optimization* **69**(3), 525–545, 2017.
- [13] M.C. Hout, M.H. Papesh, S.D. Goldinger, *Multidimensional Scaling*, *Wiley Interdisciplinary Reviews: Cognitive Science* **4**(1), 93–103, 2013.
- [14] C. Lavor, L. Liberti, N. Maculan, A. Mucherino, *The Discretizable Molecular Distance Geometry Problem*, *Computational Optimization and Applications* **52**, 115–146, 2012.
- [15] C. Lavor, L. Liberti, A. Mucherino, *The interval Branch-and-Prune Algorithm for the Discretizable Molecular Distance Geometry Problem with Inexact Distances*, *Journal of Global Optimization* **56**(3), 855–871, 2013.
- [16] C. Lavor, L. Liberti, A. Mucherino, N. Maculan, *On a Discretizable Subclass of Instances of the Molecular Distance Geometry Problem*, *ACM Conference Proceedings, 24<sup>th</sup> Annual ACM Symposium on Applied Computing*, Hawaii, USA, 804–805, 2009.
- [17] L. Liberti, C. Lavor, N. Maculan, *A Branch-and-Prune Algorithm for the Molecular Distance Geometry Problem*, *International Transactions in Operational Research* **15**, 1–17, 2008.
- [18] L. Liberti, C. Lavor, N. Maculan, A. Mucherino, *Euclidean Distance Geometry and Applications*, *SIAM Review* **56**(1), 3–69, 2014.
- [19] L. Liberti, C. Lavor, A. Mucherino, N. Maculan, *Molecular Distance Geometry Methods: from Continuous to Discrete*, *International Transactions in Operational Research* **18**(1), 33–51, 2011.
- [20] J. Mongan, C. Simmerling, J.A. McCammon, D.A. Case, A. Onufriev, *Generalized Born with a Simple, Robust Molecular Volume Correction*, *Journal of Chemical Theory and Computation* **3**, 156–169, 2007.
- [21] A. Mucherino, *On the Identification of Discretization Orders for Distance Geometry with Intervals*, *Lecture Notes in Computer Science* **8085**, F. Nielsen and F. Barbaresco (Eds.), *Proceedings of Geometric Science of Information (GSI13)*, Paris, France, 231–238, 2013.
- [22] A. Mucherino, D.S. Gonçalves, *An Approach to Dynamical Distance Geometry*, *Lecture Notes in Computer Science* **10589**, F. Nielsen, F. Barbaresco (Eds.), *Proceedings of Geometric Science of Information (GSI17)*, Paris, France, 821–829, 2017.
- [23] A. Mucherino, C. Lavor, L. Liberti, *The Discretizable Distance Geometry Problem*, *Optimization Letters* **6**(8), 1671–1686, 2012.
- [24] A. Mucherino, J.-H. Lin, D.S. Gonçalves, *Coarse-Grained Representation for Discretizable Distance Geometry with Interval Data*, to appear in *Lecture Notes in Computer Science*, 2019.
- [25] A. Mucherino, J. Omer, L. Hoyet, P. Robuffo Giordano, F. Multon, *An Application-based Characterization of Dynamical Distance Geometry Problems*, to appear in *Optimization Letters*, Springer, 2019.
- [26] J. Saxe, *Embeddability of Weighted Graphs in k-Space is Strongly NP-hard*, *Proceedings of 17<sup>th</sup> Allerton Conference in Communications, Control and Computing*, 480–489, 1979.
- [27] J. Wang, R.K. Ghosh, S.K. Das, *A Survey on Sensor Localization*, *Journal of Control Theory and Applications* **8**(1), 2–11, 2010.



# Models and Algorithms for Natural Disaster Evacuation Problems

Christian Artigues  
LAAS CNRS  
TOULOUSE, France  
Email: artigues@laas.fr

Emmanuel Hebrard  
LAAS CNRS  
TOULOUSE, France  
Email: hebrard@laas.fr

Alain Quilliot  
LIMOS CNRS UMR 6158  
LABEX IMOBS3, Université  
Clermont-Auvergne  
Bat. ISIMA, BP 10125  
Campus des Cézaux,  
63173 Aubière, France  
Email: quilliot@isima.fr

Hélène Toussaint  
LIMOS CNRS UMR 6158  
LABEX IMOBS3, CNRS  
Bat. ISIMA, BP 10125  
Campus des Cézaux,  
63173 Aubière, France  
Email: toussain@isima.fr

**Abstract**— We deal here, in the context of a H2020 project, with the design of evacuation plans in face of natural disasters: wildfire, flooding... People and goods have to be transferred from endangered places to safe places. So we schedule evacuee moves along pre-computed paths while respecting arc capacities and deadlines. We model this scheduling problem as a kind of multi-mode *Resource Constrained Project Scheduling problem (RCPSP)* and handle it through network flow techniques.

## I. INTRODUCTION

THIS work has been carried on in the context of the H2020 GEOSAFE European project [4], whose overall objective is to develop methods and tools enabling to set up an integrated decision support system to assist authorities in optimizing the resources during the response phase to a natural disaster, mainly a wildfire or a flooding. In such a circumstance, decisions which have to be taken are about fighting the cause of the disaster, adapting standard logistics (food, drinkable water, health...) to the current state of infrastructures, and evacuating endangered areas (see [2]). We focus here on the *late evacuation problem*, that means the evacuation of people and eventually critical goods which have been staying at their place as long as possible.

While evaluation planning remains mostly designed by experts, 2-step optimization approaches have been addressed [2]: the first step (pre-process) involves the identification of the routes that evacuees are going to follow; the second step, which has to be performed in real time, aims at scheduling the evacuation of estimated late evacuees along those routes. As a matter of fact, this last step involves 2 distinct work pieces, one about forecasting, difficult in the case of wildfire, because of their dependence to topography and meteorology [4], and the second one about priority rules and evacuation rates imposed to evacuees [3]. The model which we study here is closed to the one proposed in [1] and called the *non preemptive evacuation planning problem (NEPP)*. According to it, remaining evacuees have been clustered

into groups with same original location and pre-computed route, and once a group starts moving, then it must keep on at the same rate until reaching his target safe area (*Non Preemption* hypothesis, which matches practical concerns of the people who supervise the evacuation process). While authors in [1] address their model while discretizing both the time space and the rate domains and applying constraint propagation techniques, we consider it as an extension of the *Resource Constrained Project Scheduling Problem (RCPSP: [5,6])*, with continuous variables which identify evacuation rates and with an objective function which reflects the safety provided to every evacuee. We use this RCPSP reformulation in order to design a heuristic algorithm which deals with our problem according to network flow like techniques, well-fitted to real-time emergency contexts.

The paper is structured as follows: Section 2 provides the NEPP model. Section 3 describes our RCPSP reformulation. Sections 4, 5 are about algorithms and numerical tests.

## II. NON PREEMPTIVE EVACUATION PLANNING (NPEP)

We consider here a transit network  $H = (N, A)$ :  $N$  is its node set and  $A$  its arc set; Every arc  $e \in A$  is provided with the time  $TIME(e)$  required for some evacuee to move through  $e$  and with the maximum number  $CAP(e)$  of evacuees who may engage themselves  $e$  per time unit. We distinguish:

- The *Evacuation* node subset  $N^+$ , whose nodes are labelled  $i = 1..n$  and related to some population  $P(i)$ .
- The *Safe* node subset  $N$  and the *Relay* node subset  $N^-$ .

Evacuees of the population  $P(i)$  located at  $i \in N^+$  move along a pre-determined path  $I(i)$ , that means a sequence of arcs  $e^1, \dots, e^k(i)$  connecting  $i$  to some *safe* node  $S(i)$ . We set  $L\_TIME(i) = \sum_{k=1..k(i)} TIME(e^k)$ , and, for any  $k =$

1..k(i):  $L(i, k) = \sum_{k \leq j} TIME(e_k^i)$  and  $L^*(i, k) = \sum_{k \geq j} TIME(e_k^i)$ .

We must comply with capacity restrictions: During one time unit, no more than  $Deb(i)$  evacuees may start moving from  $i \in N^+$  and no more than  $CAP(e)$  evacuees may simultaneously engage themselves on a given arc  $e$ . Also, forecast about the way the natural disaster will evolve imposes that for any arc  $e$  of the transit network, nobody may start moving along  $e$  after deadline  $Dead(e)$ , while the whole evacuation process should be over at global deadline  $T-Max$ . Thus all evacuees coming from  $i \in N^+$  should reach related safe node  $S(i)$ , before  $\Delta(i) = \text{Inf}(T-Max, \text{Inf}_{k=1..k(i)}(Dead(e_k^i) + L^*(i, k)))$ .

Besides, authorities impose *Non Preemption*: once evacuees related to evacuation node  $i$  have started moving, they must keep on at the same speed and rate along path  $\Gamma(i)$ , until they all reach safe node  $S(i)$ . We denote by  $v_i$  the related evacuation rate (number of evacuees per time unit which enter on  $\Gamma(i)$  at until  $i$  becomes empty. We derive an upper bound  $v-max(i)$  for  $v_i$  by setting:  $v-max(i) = \text{Inf}(\text{Inf}_j CAP(e_k^i), Deb(i))$ . We also see that if we are provided with the start-date  $T_i$  of  $i$  evacuation process and with its evacuation rate  $v_i$  then we deduce its end-date  $T^*_i = T_i + L\_TIME(i) + P(i)/v_i$ . We deduce from deadline  $\Delta(i)$  a minimal evacuation rate  $v-min(i) = P(i)/(\Delta(i) - L\_TIME(i))$ .

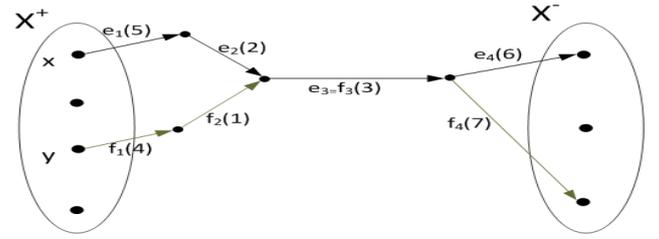
Then, the *Non Preemptive Evacuation Planning Problem* (NEPP) is about the computation of an evacuation schedule, which means of start-times  $T_i$  and evacuation rates  $v_i$ ,  $i \in N^+$ . The quality of such a schedule  $\Lambda = (T, v)$  is going to be the weighted safety margin  $\sum_i P(i) \cdot (\Delta(i) - T^*_i)$ .

### III. A RCPSP ORIENTED REFORMULATION OF NPEP.

We identify evacuation nodes  $i$  of network  $H$  and related evacuation jobs. So the key idea here is to consider the arcs  $e$  of the network  $H$  as resources, likely to be exchanged by evacuation jobs  $i, j$  whose paths  $\Gamma(i)$  and  $\Gamma(j)$  share arc  $e$ . In order to formalize it, we introduce *Conditional Time Lags*:

- If  $\Gamma(i) = \{e^i_1, \dots, e^i_{k(i)}\}$  and  $\Gamma(j) = \{f^j_1, \dots, f^j_{k(j)}\}$  share arc  $e = e^i_k = f^j_l$ , and if evacuees from  $j$  come on  $e$  after evacuees from  $i$ , then delay  $T_j - T_i$  will be no smaller than  $TL-Elem(i, j, e) = L(i, k-1) - L(j, l-1) + P(i)/v_i$ .
- Set  $Arc(i, j) = \{e \in \Gamma(i) \cap \Gamma(j)\}$  and  $TL(i, j, v_i) = \text{Sup}_{e \in Arc(i, j)} (L(i, k-1) - L(j, l-1) + P(i)/v_i) = \text{Conditional Time Lag}$  between  $i$  and  $j$ . If  $Arc(i, j) \neq \text{Nil}$  and evacuees of  $j$  enter after evacuees of  $i$  on the arcs of  $Arc(i, j)$ , then we must have  $T_j \geq T_i + TL(i, j, v_i)$ . We notice  $TL(i, j, v_i)$  depends in a convex way on the evacuation rate  $v_i$  of  $i$ .

This notion is illustrated by following Figure 1:



$$P(x) = 50, v_y = 10; P(y) = 40, v_x = 15 \Rightarrow$$

$$L(x, 2) = 7; L(y, 2) = 5; TL(x, y, v_x) = 7 + 4 - 5 = 6.$$

Figure 1: Conditional Time Lags.

We derive a RCPSP (*Resource Constrained Scheduling*: [5,6]) reformulation of NEPP, which relies on the fact that we consider every evacuation job  $i \in N^+$  as a job, whose execution requires resources which are arcs  $e \in \Gamma(i)$ , constrained by their capacities  $CAP(e)$  and whose start-dates are constrained by conditional time lags:

#### NPEP-RCPSP Model :

{*Preliminary* : We add to the set  $N^+$  two fictitious jobs  $s$  (source) and  $p$  (sink), in order to express the way resources are exchanged between jobs as a flow vector. Then we set, for any  $i \in N^+$ :  $TL(s, i, CAP(e)) = 0$  and  $TL(i, p, v_i) = L\_TIME(i) + P(i)/v_i$ .

*Output Vectors* : For any  $i$  in  $N^+ \cup \{s, p\}$  compute start-date  $T_i$  and evacuation rate  $v_i$ ; In order to do it we involve, for any pair  $(i, j)$  and any arc  $e$  in  $Arc(i, j)$  the part  $w_{i, j, e}$  of access rate to  $e$  which is given by  $i$  to  $j$

#### Constraints :

- o For any  $i \neq p$ ,  $T_i + L\_TIME(i) + P(i)/v_i \leq \Delta(i)$  ; (\*Deadline Constraints\*) (E1)
  - o for any pair  $(i, j)$  and any  $e$  in  $Arc(i, j)$ ,  $w_{i, j, e} \neq 0 \rightarrow T_j \geq T_i + TL(i, j, v_i)$ ; (\*Conditional Time Lag Constraints\*) (E2)
  - o  $T_s = 0$  ; (E3)
  - o For any  $i$  in  $N^+$ ,  $N^+$  and any arc  $e$  in  $\Gamma(i)$ , (\*Flow Constraints\*):  $\sum_j$  such that  $e \in Arc(x, y) w_{i, j, e} = v_i = \sum_j$  such that  $e \in Arc(j, i) w_{j, i, e}$ ; (E4)
  - o For any arc  $e$  of the transit network  $H$ : (\*Flow Constraints\*):  $CAP(e) = \sum_i$  such that  $e \in \Gamma(i) w_{s, i, e} = \sum_i$  such that  $e \in \Gamma(i) w_{i, p, e}$ ; (E5)
  - o For any  $i$  dans  $N^+$ ,  $v-Min(i) \leq v_i \leq v-Max(i)$ . (E6)
- Maximize:  $\sum_i P(i) \cdot (\Delta(i) - T_i - L\_TIME(i) - P(i)/v_i)$

**Explanation:** (E1) tells that every evacuation job  $i$  must be achieved before deadline  $\Delta(i)$ . (E2) means that if job  $i$  provides  $j$  with some access to arc  $e$ , then the conditional time lag inequality holds. (E4, E5) express Flow Kirshoff laws: arcs  $e$  are resources that evacuation jobs exchange between them; so job  $i$  receives  $v_i$  resource (evacuation rate) for any  $e \in \Gamma(i)$  and no more than  $CAP(e)$  such resource may be simultaneously distributed between evacuation jobs .

## IV. ALGORITHMS

NMEP model contains both NP-Hard RCPSP and TSP problems. We have to choose between assigning high rates  $v_i$  to jobs  $i$  or let them monopolize the access to transit arcs, or conversely restricting  $v_i$  in order to make  $i$  share its arcs. In order to do it, we implement a two-step approach: *MNEP-First-Step* searches a feasible schedule satisfying (E1,...,E6), while *MNEP-Second-Step* increases rates  $v_i$  in order to improve the *weighted safety margin*.

## A. The Greedy-NPEP Process.

*Greedy-NPEP* starts from some linear ordering  $\sigma$  defined on  $N^+ \cup \{s,p\}$ , and considers at any time some job  $i_0$  such that for any  $j$  prior to  $i_0$  according to  $\sigma$ ,  $v_j$ ,  $T_j$  and values  $\Pi(j,e)$  = access level to arc  $e$  that job  $j$  can transmit to  $i_0$  are available. Then it applies a 3 stage function *Assign*( $i_0$ ) which computes (see Fig. 1)  $v_{i_0}$ ,  $T_{i_0}$  and flow values  $w_{j,i_0,e}$ ,  $j$  s.t.  $j \sigma i_0$ , and  $e \in \text{Arc}(j,i_0)$ , or, in case of failure, a job  $j$ -fail  $\sigma i_0$  considered as cause of the failure.

- (1) : *Assign* scans path  $\Gamma(i_0)$ , and for any  $e$  in  $\Gamma(i_0)$ , provides  $i_0$  with access rate to  $e$  in such a way resulting *end-date*  $T^*_{i_0} \leq \Delta(i_0)$ . (see Fig. 2):

**Assign1**

For  $e$  in  $\Gamma(i_0)$  do

Let  $L\text{-Job} = \{j \text{ s.t. } (j \sigma i_0) \text{ AND } (e \in \text{Arc}(j, i_0) \text{ AND } (\Pi(j, e) \neq 0)), \text{ ordered according to increasing } T_j + TL(j, i_0, v_j) \text{ values}\}; v \leftarrow 0$ ; Not Stop;

While  $L\text{-Job} \neq \text{Nil}$  AND Not Stop do

If  $T_j + TL(j, i_0, v_j) + L\_TIME(x_0) + P(i_0)/(v + \Pi(j, e)) \leq \Delta(i_0)$  then

Compute  $w$  such that  $T_j + TL(j, i_0, v(j)) + L\_TIME(x_0) + P(i_0)/(v + w) = \Delta(i_0)$ ;

Stop;  $v \leftarrow v + w$ ;  $w_{j,i_0,e} \leftarrow w$ ;

Else  $v \leftarrow \Pi(j,e) + v$ ;  $w_{j,i_0,e} \leftarrow \Pi(j,e)$ ;

If Not Stop then Fail : Choose  $j$ -Fail in  $L\text{-Job}$

Else  $v\text{-aux}(e) \leftarrow v$ ;

If Not Fail then  $V_{i_0} \leftarrow \text{Sup}_e v\text{-aux}(e)$ ;  $e_0 \leftarrow \text{Arg Sup}$ .

$\sigma = s, \dots, x_1, \dots, x_2, \dots, x_3, \dots, x_0, \dots$

$\Gamma(x_0) = \{e_1, e_2\}$ ;  $CAP(e_1) = 20$ ,  $CAP(e_2) = 25$ ;

$\Delta(x_0)$ ;  $P(x_0) = 50$ ;  $L\_TIME(x_0) = 10$ ;  $\text{Arc}(x_2, x_0) = \{e_1\}$ ;  $\text{Arc}(x_1, x_0) = \{e_1, e_2\}$ ;  $\text{Arc}(x_3, x_0) = \{e_2\}$ ;  $TL(s, x_0) = 0$ ;  $TL(x_1, x_0) = 6$ ;  $TL(x_2, x_0) = 3$ ;  $TL(x_3, x_0) = 4$ ;

$s: T(s) = 0, \Pi(s, e_1) = 2, \Pi(s, e_2) = 3$

$x_1: T(x_1) = 5, \Pi(x_1, e_1) = 9$

$x_2: T(x_2) = 8, \Pi(x_2, e_1) = 9, P(x_2, e_2) = 14$

$x_3: T(x_3) = 13, \Pi(x_3, e_2) = 8$

$x_0$

$\Rightarrow$

*Assign-1* ->  $w_{s,x_0,e_1} = 2$ ;  $w_{s,x_0,e_2} = 3$ ;  $w_{x_1,x_0,e_1} = 8$ ;  $v_{x_0} = 10$ ; *Success*; *Assign-2* ->  $w_{x_2,x_0,e_2} = 7$ ; *Success*; *Assign-3* ->  $w_{x_1,x_0,e_1} = 0$ ;  $w_{x_2,x_0,e_1} = 8$ ;  $T_{x_0} = 21$ .

**Figure 2: Assign Process.**

- (2) : *Assign1* computes  $v_{i_0}$  and, for any  $e \neq e_0$  in  $\Gamma(i_0)$  a value  $v\text{-aux}(e)$  which may be less than  $v_{i_0}$ ; So *Assign2* increases the  $w_{j,i_0,e}$  for  $e \neq e_0$  in order to make job  $i_0$  run at the same rate for all arcs  $e$  of  $\Gamma(i_0)$ . This part of the *Assign* process may induce a failure which *Assign2* assign to some job  $j$ -Fail.
- (3) : *Assign3* makes decrease the number of arcs provided with non null  $w_{j,i_0,e}$  values by shifting values  $w_{j,i_0,e}$  which involve, for a given  $j$ , only one arc  $e$ , to another job  $j'$  such that  $e \in \text{Arc}(j', i_0)$ ,  $w_{j',i_0,e} \neq 0$  and  $\Pi(j', e) \geq w_{j,i_0,e} + w_{j',i_0,e}$ .

Then *Greedy-NPEP* comes as follows:

**Greedy-RCPSP-TL( $\sigma$ ) :**

$T_s \leftarrow 0$ ; For any arc  $e$  do  $\Pi(s, e) \leftarrow CAP(e)$ ; Not Stop;

While (Not Stop) and  $\sigma$  no fully scanned do

Apply *Assign* to current  $i_0$  and partial schedule;

If *Success*(*Assign*) then

For  $e$  in  $\Gamma(i_0)$  and  $j$  s.t.  $(j \sigma i_0) \wedge (e \in \text{Arc}(j, i_0))$

do  $\Pi(i_0, e) \leftarrow v_{i_0}$ ;  $\Pi(j, e) \leftarrow \Pi(j, e) - w_{j,i_0,e}$ ;

Else Stop; Return the pair ( $j$ -Fail,  $i_0$ ).

## B. NPEP-First-Step

*Greedy-NPEP* may fail even in the case when a solution  $(T, v, w)$  exists. It raises the question of the way we deal with linear ordering  $\sigma$ .

- Initialization of  $\sigma$** : For any  $i$ , we set  $SME(i) = \Delta(i) - L\_TIME(i) - 2.P(i)/(v\text{-max}(i) + v\text{-min}(i))$ , and compute  $\sigma$  by randomly sorting  $N^+$  in such a way that if  $P(i) < P(j)$  and  $SME(i) < SME(j)$ , then  $i \sigma j$ .
- Making  $\sigma$  evolve**. In case of failure, *Greedy-NPEP* returns a pair ( $j$ -Fail,  $i_0$ ), and this pair is inserted into a *Tabu* like set *FORBID* whose meaning is: If  $(j, i)$  is *FORBID*, then we should have  $(i \sigma j)$ .

So, global process *NPEP-First-Step* comes as follows:

**Procedure NPEP-First-Step(Max-Iter: Threshold)**

Initialize  $\sigma$  as described above; *FORBID*  $\leftarrow$  Nil;

*Iter*  $\leftarrow 0$ ; Not Stop; *Success*  $\leftarrow 0$ ;

While (*Iter*  $\leq$  *Iter-Max*) AND (Not *Success*) do

Generate  $\sigma$  consistent with *FORBID* and Apply *Greedy-NPEP*; If *Failure* then Search a failure responsible ( $j$ -Fail,  $i_0$ ) pair and put into *FORBID*.

### C. NPEP-Second-Step

In case *NPEP-First-Step* yields a feasible solution  $(T, v, w)$  *NPEP-Second-Step* improves it, by acting on rates  $v_i$  in such a way time lags  $L\_TIME(i) + P(i)/v_i$  decrease in an *ad hoc* way. Let us denote by *U-Active*, the set of pairs  $(i, j)$  which are allowed to support non null  $w_{i,j,e}$  flow values. We notice that if *U-Active* is fixed, then resulting restriction of NPEP is a convex optimization problem defined on the  $(v, w)$  polyhedron defined by (E4, E5, E6). So we fix *U-Active* according to the end of *NPEP-First-Step*, and deal with induced convex program:

- We derive from current  $v, w$ , values  $T^*_i$ , related critical paths, and values  $\lambda = \lambda(i), i \in N^+ \geq 0$ , such that  $\sum_i P(i). T^*_i = \sum_i \lambda(i)/v_i + Constant$ : Vector  $Grad = (Grad_i = -\lambda(i)/v_i^2, i \in N^+)$  is a sub-gradient vector;
- Then we modify  $v$  and  $w$  according to (I1):  $v \leftarrow v + V$ ;  $w \leftarrow w + W$ , with  $V$  and  $W$  s.t  $V.Grad < 0$  and  $v + V$  and  $w + W$  comply with (E4, E5, E6) and computed by solving *Project-Grad* following linear program:

**Project-Grad**(*U-Active*,  $v, w, \delta, Grad$ ) LP :  
 {Compute  $V = (V_i, i \in N^+)$ , and  $W = (W_{i,j,e}, (i, j) \in U-Active, e \in Arc(i, j))$  such that;  
 ○  $\forall (i, j, e), w_{i,j,e} + W_{i,j,e} \geq 0$  ;  
 ○  $\forall i \neq s, p, e \in I(i), \sum_j W_{i,j,e} = \sum_j W_{j,i,e} = V_i$  ;  
 ○  $\forall e, \sum_j W_{s,j,e} = \sum_j W_{j,p,e} = 0$  ;  
 ○  $\forall i \neq s, p, v-Min(i) \leq v_i + V_i \leq v-Max(i)$  ;  
 ○  $2.\delta \geq \sum_{i \neq s, p} V_i. Grad(i) \geq \delta$

Then *NPEP-Second-Step* comes as follows:

#### Procedure *NPEP-Second-Step*:

Let  $(T, v, w)$  be the feasible solution computed by *NPEP-First-Step* and  $T^*$  related *end-date* vector;

Derive *U-Active*; Not *Stop* ;  $Val \leftarrow \sum_i P(i). T^*_i$ ;

While Not *Stop* do

  Compute  $\delta$  and coefficients  $\lambda(i), i \in N$ ;

  Solve *Project-Grad*(*U-Active*,  $v, w, \delta, Grad$ );

  If no solution then *Stop* Else

    Apply (I1), update  $T_i, T^*_i$  and related critical paths; If  $Val-Aux = \sum_i P(i). T_i$ ; If  $Val-Aux \geq Val$  then *Stop*.

### V. NUMERICAL EXPERIMENTS.

**Purpose:** Algorithms were implemented on AMD Opteron 2.1GHz. Our goal was to evaluate the ability of *NPEP-First-Step* to deal with tight deadlines and the ability of *NPEP-Second-Step* to improve this solution.

**Instances/outputs:** An instance is a path collection  $\{I(i), i \in N^+\}$ , given together with values  $P(i), \Delta(i)$  and

$TIME(e_k^i)$ . It is summarized by a 3-uple:  $(n, m, \alpha)$ , where  $n = Car(N^+)$ ,  $m =$  number of arc  $e$ , and  $\alpha$  is as above. We both created our own instances and used an instance generator of [1]. In order to get benchmarks, we generated *ad hoc* schedules  $(T, v)$  and derived deadlines  $\Delta(i)$  which made us be provided with almost optimal solutions.

**Outputs:** For every 10 instance package, we compute:

- The number *Trial* of iterations on  $\sigma$  necessary to get a feasible solution through *NPEP-First-Step*;
- The improvement margin (%) *IMPROVE* induced by *NPEP-Second-Step*;
- The gap between *NPEP* and optimal value *VAL*

Table below provides results for  $\alpha \in [1, 2]$ .

Inst. 1: $n = 20, m = 10$	<i>Trial</i>	<i>IMPROVE</i> (%)	<i>GAP</i> (%)	<i>CPU-NPEP</i>
$\alpha = 1.2$	22.30	13.8	4.7	40.4
$\alpha = 1.5$	2.50	29.5	13.0	12.3
$\alpha = 1.7$	1.39	40.8	17.7	8.1
$\alpha = 2.0$	1.08	61.7	19.3	5.2
Inst. 1: $n = 30, m = 15$				
$\alpha = 1.2$	40.6	14.6	5.6	70.5
$\alpha = 1.5$	6.60	30.2	14.5	19.5
$\alpha = 1.7$	2.05	42.3	19.1	12.0
$\alpha = 2.0$	1.19	65.5	22.5	7.9

**Comment:** Tighting deadlines  $\Delta(i)$  improve solutions.

### VI. CONCLUSION

We described here a two-step RCPSP oriented algorithm for the *NPEP* Problem. Remains now to deal with the design of an exact method for small instances and with an integrated computation of routes  $I(i)$ .

### REFERENCES

- [1] C.Artigues, E.Hebrard, Y.Pencolé, A.Schutt, P.Stuckey: A study of evacuation planning for wildfires; *17 th Int. Workshop on Constraint Modelling/Reformulation*, Lille, France, (2018).
- [2] V.Bayram : Optimization models for large scale network evacuation planning and management : a review ; *Surveys in O.R and Management*, (2016), DOI : 10.1016/j.sorms.2016.11.001.
- [3] C.Even, V.Pillac, P.Van Hentenryk: Convergent plans for large scale evacuation; In *Proc. 29 th AAAI Conf. On Artificial Intelligence*, Austin, Texas, p 1121-1127, (2015).
- [4] Geo-Safe-; *MSCA-RISE 2015 European Project -id 691161*. <http://fseg.gre.ac.uk/fire/geo-safe.html>. Accessed Jue 12, (2018).
- [5] M.J. Orji, S. Wei. Project Scheduling Under Resource Constraints: A Recent Survey. *Inter. Journal of Engineering Research & Technology (IJERT)* Vol. 2 Issue 2, (2013)
- [6] A.Quilliot, H.Toussaint: Flow Polyedra and RCPSP, *RAIRO-RO*, 46-04, p 379-409, (2012)

# Best Response Dynamics for VLSI Physical Design Placement

Michael Rapoport  
School of Computer Science  
The Interdisciplinary Center  
Herzliya, Israel  
Email: mishkaraport@gmail.com

Tami Tamir  
School of Computer Science  
The Interdisciplinary Center  
Herzliya, Israel  
Email: tami@idc.ac.il

**Abstract**—The physical design placement problem is one of the hardest and most important problems in micro chips production. The placement defines how to place the electrical components on the chip. We consider the problem as a combinatorial optimization problem, whose instance is defined by a set of 2-dimensional rectangles, with various sizes and wire connectivity requirements. We focus on minimizing the placement area and the total wire-length.

We propose a local-search method for coping with the problem, based on natural dynamics common in game theory. Specifically, we suggest to perform variants of *Best-Response Dynamics (BRD)*. In our method, we assume that every component is controlled by a selfish agent, who aim at minimizing his individual cost, which depends on his own location and the wire-length of his connections.

We suggest several BRD methods, based on selfish migrations of a single or a cooperative of components. We performed a comprehensive experimental study on various test-benches, and compared our results with commonly known algorithms, in particular, with simulated annealing. The results show that selfish local-search, especially when applied with cooperatives of components, may be beneficial for the placement problem.

## I. INTRODUCTION

**P**HYSICAL DESIGN is a field in Electrical Engineering which deals with *very large scale integration (VLSI)*. Specifically, physical design is the main step in the creation of *Integrated Circuit (IC)*. The basic question is *how to place the electrical components on the chip*. This fundamental question became relevant with the invention of ICs in 1958 [19], and remains critical our days with the development of micro-electricity. Recent developments in micro-electricity enables transistors to reach the size of nanometers, thus a single chip can accommodate thousands of components of different sizes and dispersed connectivity. Bad layout of electrical components leads to expensive production and poor performance. Figure 1 presents the Intel i7 processor [10], and demonstrates how efficient design is crucial in enabling the accommodation of many components on a small area.

The research was supported by THE ISRAEL SCIENCE FOUNDATION (grant No. 1036/17).

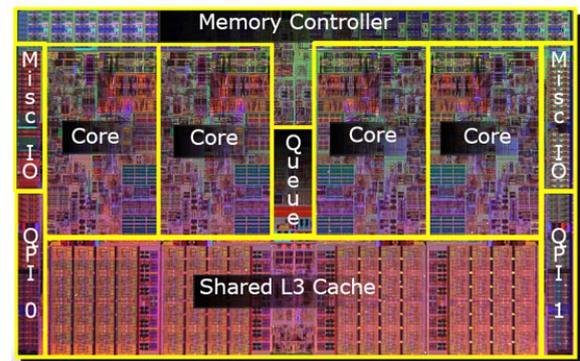


Fig. 1. Intel i7 processor placement.

The complexity of *VLSI physical design* led to the establishment of a design process, in which the problem is divided into several steps, each is an independent NP-complete problem. The most fundamental steps are: (i) *Floorplan*: choose the area of the chip and decide the positions of the building blocks of the chip, (i.e., in Intel processor: cores, graphic processor, cache, memory controller). (ii) *Placement*: Each of the above mentioned building blocks consists of several components. These components should be placed in a way that minimizes area and wire-length. (iii) *Signal and Clock Routing*: route the wires via components white space, which is an extra area assigned for wiring.

In this work, we focus on the *Placement* problem. The floorplan is usually performed manually, and the signal and clock routing is more of a production engineering problem which is tackled using different tools.

Several common methods for coping with the Placement problem are based on local-search. We propose a new class of local-search algorithms that consider the problem as a game played among the components, where each component corresponds to a selfish player who tries to maximize his own welfare. Our algorithms are different from other algorithms based on local search in the way they explore the solution space. Every solution is associated with a global cost, and every component is associated with its individual cost, which is based on its own placement and connections. We move from one solution to another if this move is selfishly beneficial for

a single component or for a small cooperative of components, without considering the effect on the global cost.

In this paper we first review the placement problem, and survey some of the existing techniques to tackle it, in particular local-search algorithms and Simulated Annealing. We then describe our new method of performing selfish Best-Response Dynamics (BRD). In order to evaluate this method, we performed an extensive experimental study in which we simulated and tested several variants of BRD on various test-benches. Our results show that BRD may produce a quick and high quality solution.

#### A. The Placement Problem

The Placement process determines the location of the various circuit components within the chip's core area. The problem, and even simple subclasses of it, were shown to be NP-complete by reductions to the *bin packing* and the *rectangle packing* problems [14], [15]. Moreover, a reduction to the *Quadratic assignment problem* shows that achieving even a constant approximation is NP-hard [9].

Bad placement not only reduces the chip's performance, but might also make it non-manufacturable by forcing very high wire-length, lack of space, or failing timing/power constraints. As demonstrated in Figure 2, good placement involves an optimization of several objectives that together ensure the circuit meets its performance demand [4], [17]:

- 1) *Area*: Minimizing the total area used to accommodate the components reduces the cost of the chip and is crucial for the production.
- 2) *Total wire-length*: Minimizing the total length of the wires connecting the components is the primary objective of the physical design. Long wires require the insertion of additional buffering, to insure synchronization between the components. Short wires decrease the power consumption and the system's leakage.
- 3) *Wire intersection*: Our days, wire intersection is allowed as long as a single wire does not have more than a predefined number of intersections. The manufacturing process enables several routing layers. Nevertheless, a good layout avoids unnecessary intersections.
- 4) *Timing*: The timing cycle of a chip (clock frequency) is determined by the delay induced by the longest wire, usually referred to as the critical path.

Our work considers the initial placement calculation, denoted *global placement*. This stage is followed by the *detailed placement* stage, in which the global placement results are put into use and the cells are actually placed on the die. The detailed placement stage includes small changes to solve local issues such as wire congestion spots, remaining overlaps, layout constrains (such as via locations), connecting to the die pinout, etc.

In the global placement stage, several parameters are optimized. We focus on the total wire-length and placement area. By adjusting the cost function associated with each configuration, our method enables considering additional parameters such as wire congestion, critical path length, and more.

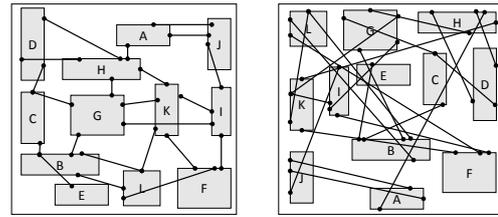


Fig. 2. An example of a good placement (left) v.s a bad placement (right). In the good placement the wires are shorter and there is almost no congestion. In this example, the area of both placements is the same.

#### B. Formal Description of the Placement Problem

We describe the placement problem as a combinatorial optimization problem. The components composing the problem are represented by 2-dimensional rectangles denoted *blocks*. In the placement, they can be rotated by  $90^\circ$ ,  $180^\circ$  or  $270^\circ$ , but not mirrored. The sides of the assigned blocks must be parallel to each other, and to the bounding area. Whenever we refer to a location in a block, we let  $(0, 0)$  be the bottom-left corner, and every other point in the block is given by its  $(x, y)$  coordinates with respect to this corner.

Formally, an instance of the problem is defined by:

- 1) A set of  $n$  blocks  $\{B_1, B_2, \dots, B_n\}$  to be placed on the chip. Every block  $1 \leq i \leq n$ , has associated width  $w_i$  and height  $h_i$ .
- 2) A list of required connections between the blocks,  $\{N_1, N_2, \dots, N_m\}$ . Every connection is given by a pair of blocks, and the locations in which these blocks should be connected. Formally  $N_j = \langle B_j^1, x_j^1, y_j^1, B_j^2, x_j^2, y_j^2 \rangle$ , for  $0 \leq x_j^1 \leq w_j^1, 0 \leq y_j^1 \leq h_j^1$  and  $0 \leq x_j^2 \leq w_j^2, 0 \leq y_j^2 \leq h_j^2$ , corresponds to a request to connect blocks  $B_j^1$  and  $B_j^2$ , such that the wire is connected to coordinate  $(x_j^1, y_j^1)$  in  $B_j^1$  and to coordinate  $(x_j^2, y_j^2)$  in  $B_j^2$ .

The output of the problem is a placement  $F$  given by the locations of the blocks on the plane  $\{L_1, L_2, \dots, L_n\}$ , such that for every  $1 \leq i \leq n$ ,  $L_i = (x_i, y_i, r_i)$ . The parameter  $r_i \in \{0, 1, 2, 3\}$  specifies how block  $B_i$  is rotated corresponding to  $\{0, 90, 180, 270\}$  degrees. Note that a rotation by  $180^\circ$  is not equivalent to not rotating at all, since the location of the required connections is also rotated. Formally, block  $B_i$  is placed in the rectangle whose diagonal endpoints are  $(x_i, y_i)$  (this corner is independent of the value of  $r_i$ ), and  $(x_i + w_i, y_i + h_i)$  if  $r_i = 0$ , or  $(x_i + h_i, y_i + w_i)$  if  $r_i = 1$ , or  $(x_i - w_i, y_i - h_i)$  if  $r_i = 2$ , or  $(x_i - h_i, y_i + w_i)$  if  $r_i = 3$ .

A placement is legal if no two blocks overlap, that is, the rectangles induced by  $L_{i_1}$  and  $L_{i_2}$  are disjoint for all  $i_1 \neq i_2$ .

This condition may be relaxed a bit in the global placement stage, and allow small percentage of overlaps area. These overlaps are resolved later during the detailed placement stage.

The *bounding box* of a Placement  $F$ , is the minimum axis-aligned rectangle which contains all the blocks. The *area* of a placement  $F$  is the area of the bounding box, and is denoted  $A(F)$ .

The blocks' location, together with the required connections, induce the wire-length of a placement. Formally, assume

that blocks  $B_1$  and  $B_2$  are located in  $L_1$  and  $L_2$ , respectively, and let  $N_j = \langle B_j^1, x_j^1, y_j^1, B_j^2, x_j^2, y_j^2 \rangle$ . We first calculate the actual coordinates of the connection points, based on  $L_1, L_2$ , and the values of  $\langle x_j^1, y_j^1 \rangle$  and  $\langle x_j^2, y_j^2 \rangle$ . Let  $\langle \hat{x}_j^1, \hat{y}_j^1 \rangle$  and  $\langle \hat{x}_j^2, \hat{y}_j^2 \rangle$  be the points we need to connect. The wire-length associated with  $N_j$ , denoted  $Len(N_j)$ , is calculated in a way that fits the actual production process, in which all wires are parallel to the blocks and to the bounding area, that is,  $Len(N_j) = \Delta X + \Delta Y = |\hat{x}_j^1 - \hat{x}_j^2| + |\hat{y}_j^1 - \hat{y}_j^2|$ . The total wire-length of a Placement  $F$  is denoted  $L(F)$ , and is given by  $L(F) = \sum_{j=1}^m Len(N_j(F))$ . An example of wire-length calculation is given in Figure 3.

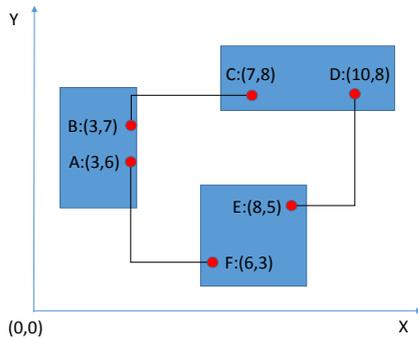


Fig. 3. An Example of wire-length calculation. There are three connections between the pairs of points  $\{A, F\}$ ,  $\{D, E\}$  and  $\{B, C\}$ . The total wire-length is  $(|X_A - X_F| + |Y_A - Y_F|) + (|X_E - X_D| + |Y_E - Y_D|) + (|X_B - X_C| + |Y_B - Y_C|) = (3 + 3) + (2 + 3) + (4 + 1) = 16$ .

The goal in the placement problem is to minimize  $\alpha L(F) + (1 - \alpha)A(F)$  where the parameter  $0 \leq \alpha \leq 1$  weight the importance of the two objectives. These days (as the number of components per chip rises) it is a very common practice to focus on the wire-length of the placement and only when finished optimizing the wire-length, perform small changes in order to gain better area result, with a minimal harm of the achieved wire-length. Thus, in our experiments (to be described in Section III), we give a substantially higher weight to the wire-length.

### C. Current Techniques for Efficient Placement

We now overview the common disciplines to handle the Placement problem. Some algorithms are tailored for simplified classes of instances. Specifically,

- 1) *Standard cell*: Components may have different width, but they all have the same height and are placed in rows. With over-cell routing the goal is to minimize the width of the widest row and the total wire-length.
- 2) *Gate array / FPGA*: The area is discretized to equally sized squares where each square is a possible component location. All the components have the same size and shape but different connections between them, the goal is to minimize the total wire-length.

Both classes induce simplified problems, which are still NP-hard, but can be approximately solved using Linear Programming [22], [7], Greedy Algorithms [25], [2], Slicing Tree

representation [3], or by Divide and Conquer algorithms that allows temporal block overlaps [2], [7].

A different solution approach is to develop heuristics, usually with strong randomness involved. Most heuristics have no assumptions on the problem thus able coping with general instances. Heuristics have no performance guarantee but perform well in practice. Some heuristics were tailored for *Standard cell* and *Gate array* instances [8], [16], [24]. The most commonly used algorithm concept for placement is *simulated annealing* (SA) [23], [20]. Modern algorithms of our days are always compared against it and many of them are based on its concept. While SA is unlikely to find an optimal solution, it can often find a very good one. The name simulated annealing come from annealing in metallurgy, a technique involving heating and controlled cooling of a material to increase the size of its crystals and reduce their defects. Both are attributes of the material that depend on its thermodynamic free energy. Heating and cooling the material affects both the temperature and the thermodynamic free energy. The simulation of annealing as an approach for minimization of a function of large number of variables was first formulated in [13]. Many modern algorithms are based on the concepts of simulated annealing.

Additional widely used placement methods include (i) Force Directed Placement, in which the problem is transformed into a classical mechanics problem of a system of objects attached to springs [21], (ii) Placement by Partitioning, in which the circuit is recursively partitioned into smaller groups [2], [7], (iii) Numerical Optimization Techniques, based on equation solving and eigenvalue calculations [25], [18], and (iv) Placement by Genetic Algorithm, that emulates the natural process of evolution as a means of progressing toward optimum, [25]. Some of these methods are only suited for *Standard cell* or *Gate array* instances, and some are general. A survey of the above and of additional algorithms for placement can be found in [1], [24], [11].

## II. OUR LOCAL-SEARCH METHOD FOR SOLVING THE PLACEMENT PROBLEM

The main challenges involved in solving the Placement problem are the need to optimize several objectives simultaneously, and to achieve even a good approximate solution in reasonable time. Optimizing even a single objective is an NP-hard problem. Naturally, combining several objectives, that may be conflicting, makes the problem more challenging.

Our proposed method not only performs a good placement in a relatively short time, but also copes with the multiple objective challenge.

### A. The Placement Problem as a Game

We propose to tackle the problem by a local-search algorithm, using natural dynamics common in *game theory*. Specifically, we suggest to perform variants of *Best-Response Dynamics* (BRD), assuming the components correspond to strategic selfish agents who strive to optimize their own welfare. In a BRD process, every agent (player) in turn, selects

his best strategy given the strategies of the other players. In our game, the strategy space of a player consists of all the locations his component can be placed in, given the location of the other components. Players keep changing strategies until a Nash equilibrium of the game is reached. A Nash equilibrium is a strategy profile in which no player can benefit from changing his strategy [12]. A lot of attention is given to best-response dynamics in the analysis of non-cooperative games, as this is the natural method by which players proceed toward a NE. The common research questions are whether BRD converges to a NE, the convergence time, and the quality of the solution (e.g. [5], [6]).

BRD can also be performed with coordinated deviations. That is, in each step, a group of players, denoted a *cooperative*, moves simultaneously, such that their total cost is reduced. Note that in a coordinated deviation of a *cooperative*, unlike a *coalition*, some members of the cooperative may be hurt. The deviation is beneficial if the total members' cost is reduced.

In order to consider the placement problem as a game played by selfish agents, we need to associate a value, or cost, for each player in each possible configuration of the game. In our setting, players correspond to blocks and configurations correspond to placements. The BRD process is defined with respect to a cost function that depends on the wire-length connected to the player's block, and the total placement area. The individual cost function is calculated for each block or cooperative, and is relevant only to the currently playing block or cooperative.

Recall that for a configuration  $F$ , the global cost of  $F$  is

$$Global\_cost(F) = \alpha L(F) + (1 - \alpha)A(F),$$

where  $L(F)$  is the total wire-length,  $A(F)$  is the bounding box area, and the parameter  $\alpha$  is used to weight these two components of the cost function. In our algorithms, the global cost function, is used only to evaluate the final configuration - in order to compare different methods and to analyze the progress of the algorithms.

The *individual cost function* is used to evaluate the possible deviations of the currently playing block. For a single block  $B_i$ , let  $L_{B_i}(F)$  denote the total wire-length of  $B_i$ 's connections. By definition,  $L(F) = \frac{1}{2} \sum_{1 \leq i \leq n} L_{B_i}(F)$ . The total individual wire-length is divided by 2, since every wire is counted in the individual wire-length of its two endpoints. For a configuration  $F$  and a block  $B_i$ , the individual cost of  $B_i$  in  $F$  is defined as follows:

$$Ind\_cost(B_i, F) = \alpha \cdot (L_{B_i}(F))^2 + (1 - \alpha) \cdot A(F).$$

Note that in the individual cost function, the corresponding wire-length is squared - for normalization with the area component.

Let  $\Gamma$  be a subset of the blocks. In order to evaluate configurations that are a result of a coordinated deviation, we define, for a cooperative  $\Gamma$  in a configuration  $F$ , the individual cost of  $\Gamma$  in  $F$ :

$$Ind\_cost(\Gamma, F) = \alpha \cdot \sum_{B_i \in \Gamma} (L_{B_i}(F))^2 + (1 - \alpha) \cdot A(F).$$

Since finding the best response is NP-hard in most scenarios and particularly for coordinated deviation, we perform a better-response move, in which the player (or a cooperative) benefits, but not necessarily in the optimal way. In practice, we perform the best response move in a restricted search space. Also, in some algorithms, when there is no local improving step, we may perform a move which harms the cost function. Such moves result in a temporary worse state and are used in order to allow the algorithm to escape from local minima.

In our experiments, we compared our results with those achieved by *Simulated Annealing* (SA), *Greedy Local Search* (GLS) algorithm, based on hill climbing, and *Fast Local Search* (FLS) algorithm (faster version of the greedy local search). For each test-bench we run our algorithms as well as these algorithms, and compared the results. In this paper we only provide the comparison with SA, as it outperformed the other two local-search methods.

### B. Search for a Solution over the Solution Space

Before presenting our algorithms we give an overview of the local search technique, and explain how the search for the solution is performed. A local-search algorithm performs a search over the solution space. Every possible solution (placement  $F$ ) is associated with a score ( $Global\_cost(F)$ ). The global cost function defines a placement-cost multidimensional complex, on which the algorithm advances. Each point on the complex is a placement and the complex includes all possible placements.

Every local-search algorithm moves on the placement-cost complex searching for a point corresponding to a placement having minimum cost. The local-search paradigm implies that the movement along the complex is almost continuous. When the algorithm encounters a heap on the complex which it cannot pass, it may try to bypass it in order to continue the search in that direction.

The main challenge when applying such algorithms, is how to pick the next point to explore and how to decide when to stop the search. As demonstrated in Figure 4, we can continue to search up to some point of worse cost but we do not know what awaits us further down the path of the search. We may attempt to remember each minimum we visit during the search and traverse different search paths from each local minimum detected. However, such methods perform a brute force search, which in turn results in exponential running time. Finding the global minimum means we have found an optimal solution for an NP-hard problem. Hence, such algorithms must have exponential running time (regardless of the algorithm's logic) unless P = NP.

The main difference between our algorithms and previous algorithms based on local search (in particular SA, GLS, FLS), is the way we evaluate each solution in the solution space, and the way we advance to the next solution in the search process. Previous algorithms calculate the cost function for the entire placement, while our algorithms base their progress on the individual cost of a block (or of a cooperative of blocks). The main goal of this work is to examine the quality of local-search algorithms for the placement problem, in which

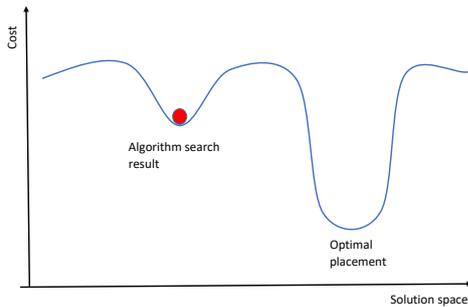


Fig. 4. A general description of a local-search method. The algorithm traverses the cost-placement multidimensional complex. Local search may end-up in a local minimum, unable to escape, thus also unable to find an optimal placement.

the search on the solution space is determined by only a single block (or a small cooperative), in a selfish manner, according to its individual cost-over-time curve. Our method does not use a global cost function; we present the global objective parameters (wire-length, total area, etc.) only for the analysis. As detailed in Section II-C, some of our algorithms accept moves that harm a bit the individual cost function (thus breaking the selfishness to some extent). This possibility allows the algorithms to escape local minimum and the search becomes much more versatile.

### C. Algorithms Based on Best-Response Dynamics

The BRD process proceeds as follows: Every block corresponds to a player. In every step a player or a cooperative of players have a chance to change their current location, in a way that reduces their individual cost in the resulting configuration. In some of our algorithms, a step increasing the individual cost may be accepted with some probability.

Every block can perform one or more of the following moves: {Up, Down, Left, Right, Rotate 90°, Rotate 180°, Rotate 270°}. A step is legal as long as the resulting configuration is legal.

We use three search methods for the block migrations:

1. *Best Response (BRD)*: Each block is controlled by a selfish player. Each player can perform one move per turn, the move is the best local move the block can perform to reduce its individual cost. The algorithm advances in rounds, where in each round, every player gets an opportunity to migrate. Players are allowed to perform only legal moves (no block-overlaps are created).

2. *Constant Perturbations (BRD-ConstPerb)*: In This variant of BRD, when a player does not have a legal improving move to perform, he may, with some non-negligible probability, choose a step which harms its individual cost. In our experiments we found 0.3 to be a good probability for accepting a worse state. It is small enough not to harm the selfishness on one hand, and allows the placement to escape local minima on the other.

3. *Relaxed Search (BRD-RlxSrch)*: This algorithm is another variant of BRD. The difference is that players can select illegal

locations - that involve block overlaps. While blocks are not allowed to overlap beyond a reasonable limit in the final configuration, temporal overlaps may be fruitful. Our relaxed search allows overlaps with varying fines on the area of the overlap. The overlaps fine are added to the block's individual cost. The fines are increased every round - to encourage convergence to a final placement with hardly any overlaps. The Global placement stage can tolerate small overlaps, so the output is accepted if the final placement does include some overlaps.

Each of the above algorithms is ran in two variations: without and with *swap moves*. A swap move is a move in which the active block swaps places with some other block if the swap is legal and reduces the active block's individual cost, as well as the global cost function (this ensures we avoid recurrent swaps between a pair of blocks). Swap moves break the locality of the search and allows another method with which to escape local minima. Instead of attempting to escape a local minimum by accepting a worse state, the algorithm can escape a local minimum by jumping to a better, yet not local neighbor, state. Swap moves do not break the selfishness of our algorithms but rather only the locality, and only to some extent. As our experiments reveal, enabling swap moves improves the quality of the solution.

### D. Coordinated Deviation of a Cooperative

Unlike a unilateral deviation, a coordinated deviation is initiated by a group of players, denoted a *cooperative*, who migrate simultaneously. Such a migration may harm the individual cost of some cooperative members (for example, if they give up good spots for other members), however, the total cost of the cooperative members is strictly reduced. When applying a coordinated deviation, we first determine the cooperative size and then the blocks composing it. A coordinated deviation is therefore defined by (i) the search method, (ii) the cooperative size method, and (iii) the cooperative member selection method.

We simulated three different methods for determining the cooperative's size:

- 1) *Increasing*: Starting from  $k = 1$ , after converging to a  $k$ -NE profile, which is stable against deviations of cooperatives of size  $k$ , we increase the active cooperative size to  $k + 1$ . We keep increasing the cooperative size up to a predefined limit.
- 2) *Iterating*: Each round has a different cooperative size, the sizes are incremented after each round, when the size reaches a predefined limit we reset the size to a single block.
- 3) *Random*: Each cooperative has a random size, the size is uniformly distributed between a single block and a predefined limit.

The cooperative's members are selected in the following way: we iterate over all the blocks, selecting a different *head block* of the cooperative in each round. The head block constructs a cooperative according to one of the following methods:

- 1) *Closest Connected blocks*: in every iteration we add to the cooperative a block with the shortest wire-length to some other block already in the cooperative.
- 2) *Farthest Connected blocks*: in every iteration we add to the cooperative a block with the longest wire-length to some other block already in the cooperative.
- 3) *Closest Geometrically blocks*: in every iteration we add to the cooperative a block with the smallest closest geometrical distance to the head block of the cooperative.
- 4) *Farthest Geometrically blocks*: in every iteration we add to the cooperative a block with the highest geometrical distance to the head block of the cooperative.
- 5) *Random*: Random set of blocks. The cooperative is built by uniformly adding blocks one by one, until the cooperative size is reached.

In our experiments, we run and compared various combinations of search algorithms with cooperative size and formation methods. The algorithms advance as follows: once the cooperative has been formed, all the feasible permutations of possible moves for the cooperative (depending on the search algorithm) are calculated. For each permutation we calculate the individual cost of the cooperative in the resulting placement. The permutation that minimizes this cost is chosen. Only the cooperative cost is taken into account, and we ignore the global cost as well as the internal distribution of the cost among the blocks composing it.

### E. Expected Algorithms' Progress

In this section we review our algorithms by describing their progress in general. A typical cost-over-time progress of BRD-ConstPerb is depicted in Fig 5. Since players can choose a step which harms their individual cost, we expect the algorithm to be able to escape local minima by moving to a more expensive placement and improving it by a sequence of cost-reducing moves, which hopefully lead to a better local minimum.

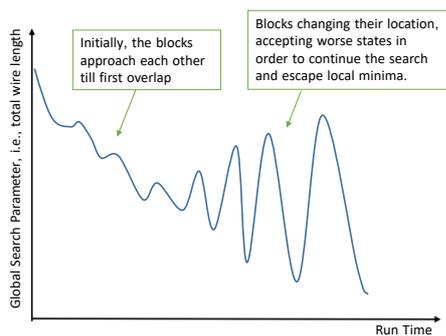


Fig. 5. Expected progress of the *BRD-ConstPerb* search method with unilateral deviations.

The progress of our BRD-RlxSrch method depends heavily on the fines for overlaps. Recall that these fines increase with the run time. As illustrated in Figure 6, this enables the algorithm to explore more areas in the solution space. Once the fines are above some threshold, the algorithm explores

feasible or almost feasible solutions whose cost may be higher than former non-feasible solutions explored earlier.

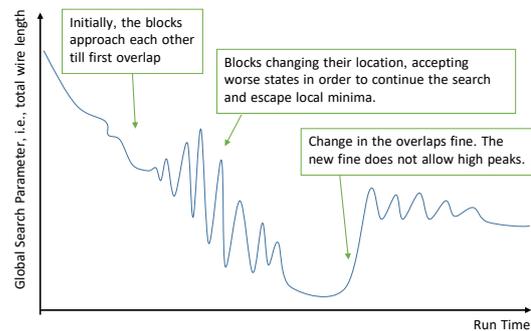


Fig. 6. Expected progress of the *BRD-RlxSrch* method with unilateral deviations. The slope has peaks, the cost function is monotonically decreasing via game of tradeoffs between the search parameters and the overlaps.

Finally, Figure 7 illustrates the typical cost-over-time progress of BRD-RlxSrch and BRD-ConstPerb when coordinated deviations are allowed. The possibility to accept worse or unfeasible solutions enables the algorithm to escape local minima and to advance in the search space towards a better solution.

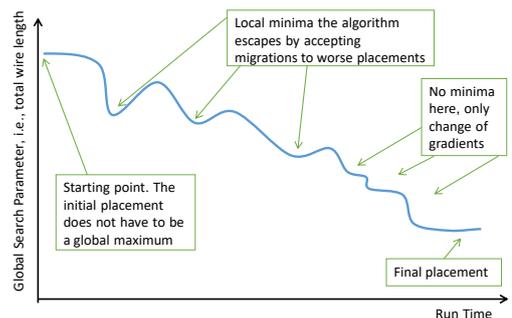


Fig. 7. Expected progress of *BRD-ConstPerb* and *BRD-RlxSrch* with coordinated deviation.

In the above figures we present the slopes as monotonically nicely curved lines, in reality this is not the case. The real lines have various gradient changes, and they are far from being nicely curved over the monotonic movement sections. The figures present the tendency of the algorithm and the overall progress.

## III. EXPERIMENTAL RESULTS

In this section we present our experimental study. Our experiments simulate the global placement stage. This stage is followed by the detailed placement stage, in which the global placement results are applied and the blocks are actually placed on the die. Usually at the detailed placement stage, small changes occur in order to solve some local issues such as wire congestion spots, remaining overlaps, layout constraints, connecting to the die pinout, etc.

We first demonstrate our concept by presenting the results of the unilateral deviation algorithms. Next, we compare some of

our heuristics with the *Simulated annealing* algorithm. Finally, we study coordinated deviations and analyze the effects of the cooperative size and structure. We explore how coordinated deviations improves the results obtained by unilateral deviations, regardless of the selected method for the search algorithm, and also consider algorithms that combine unilateral and coordinated deviations.

#### A. Experiments Setup

All algorithms are ran on the same machine with similar conditions. We sample various parameters during the algorithms run, in order to study not only the final outcome but also the search process. Time measurement is conducted and counted by the algorithms context timers, thus if a context switch occurs the timer pauses. While the time values themselves vary on different machines, the progress of the algorithms and comparison between them is valid and independent of the machine.

Our experiments were performed on 6 different test-benches,  $T_{30}^4, T_{30}^6, T_{30}^8, T_{40}^4, T_{40}^6$  and  $T_{40}^8$ , where  $T_n^c$  corresponds to an instance of  $n$  blocks, with  $c$  connections-per-block. In all instances, the block sizes are randomly distributed, height and width being a random equally distributed number in the range  $[30, 80]$  (pixels). The different connections-per-block parameter enables a good comparison and allows us to isolate and emphasize various aspects of the algorithms.

Recall that the Individual cost of a block  $B_i$  in a placement  $F$  is defined to be  $Ind\_cost(B_i, F) = \alpha \cdot (L_{B_i}(F))^2 + (1 - \alpha) \cdot A(F)$ . We run the search algorithms with various values for  $\alpha$ , and found out that the wire-length component should get much higher weight. Thus, all the results described in this section were obtained with  $\alpha = 0.9$ . This fits the common practice these days to focus on minimizing the wire-length of the placement and only when done optimizing the wire-length, perform small changes in order to gain better area result.

As detailed in Section II-C, we used three local search method: BRD – only legal profitable moves, BRD-ConstPerb – legal but maybe harmful moves, and BRD-RlxSrch – profitable but maybe non-legal moves (overlaps associated with fine). These search methods are constructed into algorithms, combining unilateral players and coordinated deviation players.

Each of these local search methods run in two different variations, without or with *swap moves*. Note that a swap move differs from a cooperative of size 2. The two members of a cooperative may swap places as long as it improves their total cost. However, a swap move is initiated by a single block and may hurt significantly the individual cost of the second block involved.

In order to better evaluate the algorithms, we performed each experiment several times. Specifically, we ran the algorithms on the same instance with 5 different random initial placements. While the initial placement has a strong impact on the results, the final result depends on the progress of the algorithm as much as on the initial placement. If for any test bench one algorithm is better than the other, then it is almost

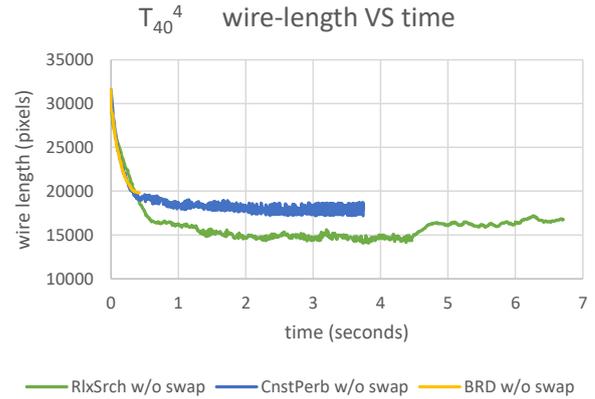


Fig. 8. Progress of total wire-length. No swap moves.

certain better for any initial placement. The variation between the results of different algorithms is consistent for most of the initial placements. In order to compare the algorithms we look at the average results over all initial placements.

#### B. Results for Unilateral Deviations

The first experiment we present is a comparison of the three search methods, when applied without and with swap moves. We run each of these variants on our test-bench  $T_{40}^4$ , that is, an instance consisting of 40 blocks each with 4 connections. All the algorithms were applied starting from the same initial placement. Figures 8 and 9 present the progress of wire-length over running-time without and with swaps, respectively. Figures 10 and 11 present the progress of area over running time with and without swaps, respectively. We can see that the algorithms reach their stopping criteria at different times and have different progress gradient.

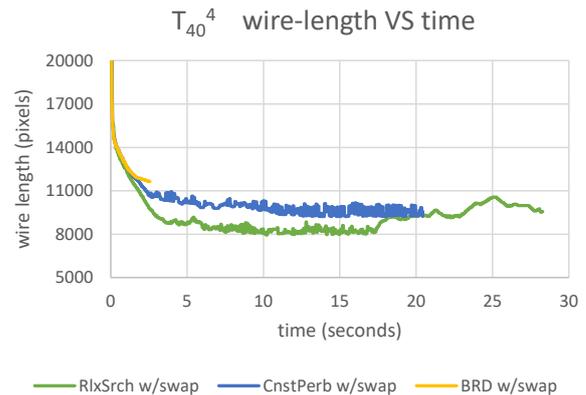


Fig. 9. Progress of total wire-length. Swap moves allowed.

The BRD algorithm is the first to finish - its local search is more restricted and thus, the stopping criteria is reached earlier. Each algorithm has a different progress curve. The gradient of the changes in the cost function according to the time is different. Nevertheless an obvious pattern can be observed: the algorithms that can progress to a worse state

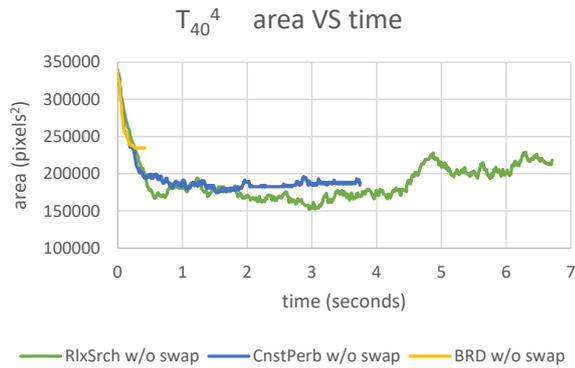


Fig. 10. Progress of Bounding-box area. No swap moves.

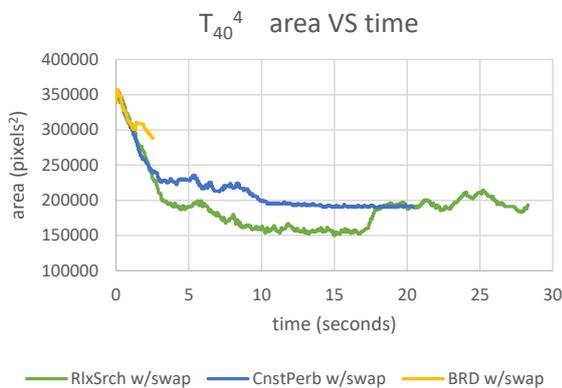


Fig. 11. Progress of Bounding-box area. No swap moves.

tend to continue and improve their result as the algorithm progresses. Such algorithms are able to escape local minima and continue the search, thus obviously the progress gradient is much more moderate. Moreover, we also witness the major influence of swaps. Allowing swaps increases the running time but improves the result. Such behavior leads to having a more moderate gradient of change, but due to the increase in run time, we eventually reach a lower level and a better result.

For BRD-RlxSrch and BRD-ConstPerb we can see peaks and drops in the performance – corresponding to reaching and escaping local minima. In BRD the spikes are limited due to the search method, which always chooses an improving step. Still, the curve is not monotonically decreasing as the improvement are with respect to the deviating block’s individual cost, that may conflict with the global cost.

Due to space constraints we do not present the plot presenting the progress of the overlap area in the BRD-RlxSrch algorithm. In both applications, with or without swaps, the overlap area is not increasing or decreasing monotonically. Initially, the algorithm explores non-feasible solutions, that have low wire-length and bounding box area; however, as the fine for overlaps is increased, the placements become more and more overlap-free. The final placement achieved by the relaxed search algorithm is feasible, and its quality is more or less equivalent to the quality achieved by BRD-ConstPerb.

### C. Comparison with the Simulated Annealing Algorithm

In this section we present a comparison of our unilateral deviation algorithms with the Simulated Annealing (SA) algorithm. We present the results by normalizing the SA result to 1. We ran SA and each of our algorithm on all six test-benches, and normalized the result with respect to the corresponding SA result. For example, if on some instance the SA algorithm produces a placement whose area is 12000 pixel<sup>2</sup>, and our algorithm produces a placement whose area is 9600 pixel<sup>2</sup>, then the result of our algorithm is presented as 9600/12000 = 0.8.

When presenting the results, we distinguish between two groups of algorithms. The first group includes algorithms that are more run-time oriented than result-oriented, while the other aim to achieve a good result. The first group consists of BRD with and without swaps, and BRD-ConstPerb without swaps, while the second group consists of BRD-RlxSrch with and without swaps, and BRD-ConstPerb with swaps.

Figures 12 and 13 present the results for the total wire-length, and Figures 14 and 15 present the results for the placement area. In all the figures, the horizontal black line represents the result achieved by the SA algorithm.

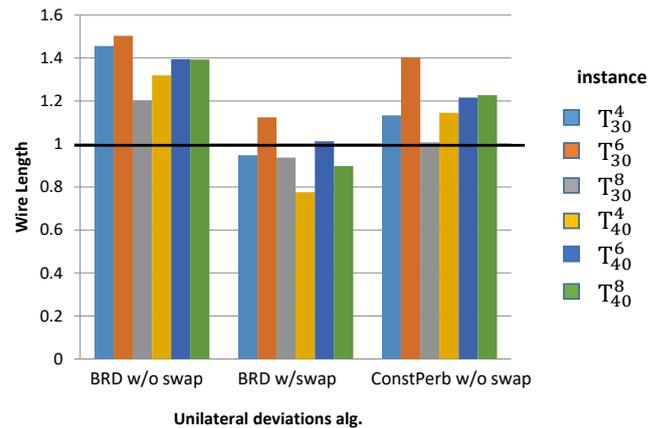


Fig. 12. Wire-length normalized to SA performance – Group I

Figures 16 and 17 present the comparison between the run-times of the algorithms. The algorithm of the first group are indeed much faster than SA. Also, all algorithms when run without swaps are at least 5 times faster than SA.

The experiments reveal that we can achieve better results of both wire-length and placement area while paying with a slightly worse running time. As well as the other way around, that is, slightly worse result can be achieved with a fraction of the running time. We also witness certain algorithms, which on the majority of the test benches, have succeeded to achieve a better result in a lower running time.

### D. Coordinated Deviation

As detailed in Section II-D, coordinated deviations are performed by a cooperative of blocks. In this section we analyze the results achieved by our search algorithms when

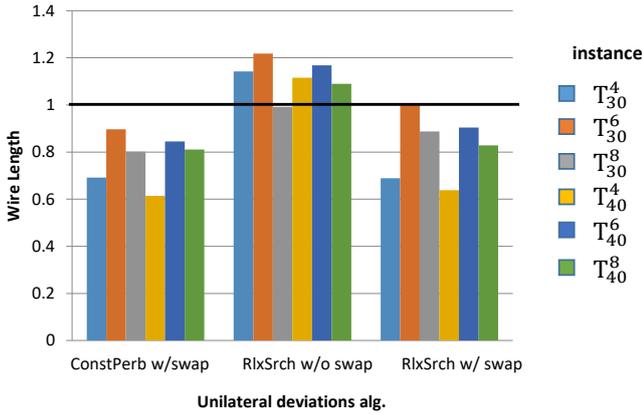


Fig. 13. Wire-length normalized to SA performance – Group II

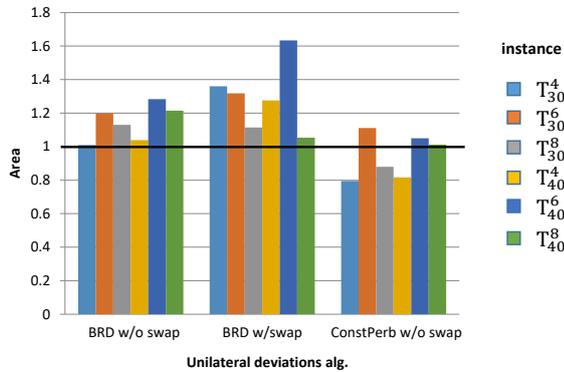


Fig. 14. Area normalized to SA performance – Group I

applied with coordinated deviations. Recall that a deviation of a cooperative is beneficial if the total cost of the cooperative members is strictly reduced. In addition to the local-search method (BRD, BRD-ConstPerb and BRD-RlxSrch), the algorithms are different in the way they determine the active cooperative size and formation. In all the experiments with coordinated deviations, the algorithms were performed on the same instance and the same initial placements.

Due to space constraints, we do not present the results

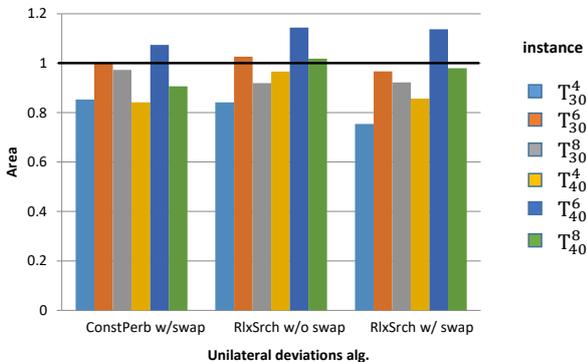


Fig. 15. Area normalized to SA performance – Group II

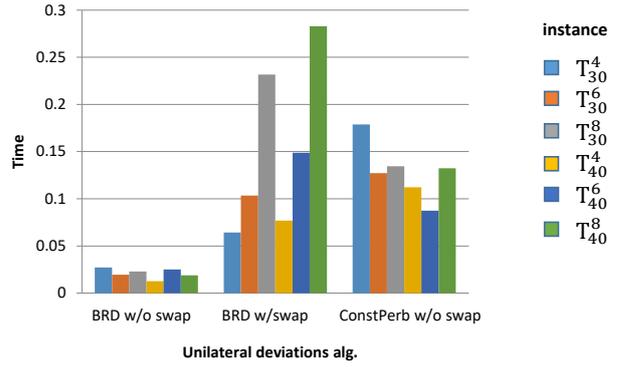


Fig. 16. Running time normalized to SA performance – Group I. The line corresponding to SA (Time = 1) is not shown, as it is way above the shown bars.

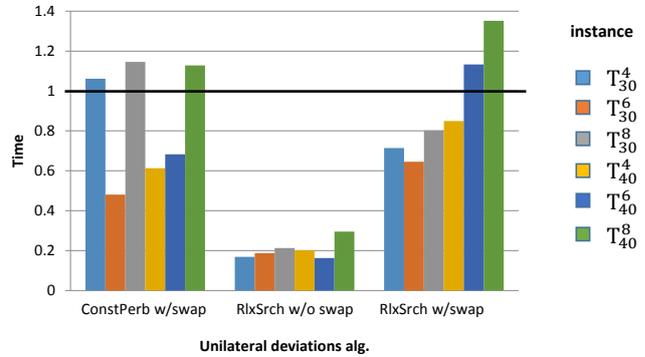


Fig. 17. Running time normalized to SA performance – Group II

using plots, and only summarize our conclusions. Our results show that the cooperative size has the largest influence on the results. The method for determining the cooperative size in each round is not crucial as the predefined limit for the maximal cooperative size. The higher this limit, the better the results. In addition, the experiments do not crown a specific method for selecting the cooperative members - the results vary and depend heavily on the initial placement.

We witnessed a major improvement in the results already with cooperatives of two blocks (compared with unilateral deviations). Further increase in the cooperative size do tend to improve the result, however it involves an exponential increase in the running time. Therefore, the best is to run the algorithm initially with cooperatives of size two, and allow non-frequent rounds in which larger cooperatives are activated. Such executions converge to a final placement much faster than SA, and if performed multiple times, with different initial placements, are expected to produce at least one excellent outcome.

We believe that this unique algorithm, that combines our search methods with a mixture of coordinated and unilateral deviations, is the main result of this work.

#### IV. SUMMARY AND CONCLUSIONS

In this work we examined the performance of local search algorithms for the global placement problem in VLSI physical design. Our algorithms are different from common local search algorithms in the way they explore the solution space. Every solution is associated with a global cost (based on its bounding-box area, the total wire-length, and possibly additional parameters), and every component is associated with its individual cost (based on its own placement and connections). We explore the solution space by moving from one solution to another if this move is selfishly beneficial for a single or for a cooperative of components, without considering the effect on the global cost. Best-response dynamics (BRD) is performed until no component has a beneficial migration. We suggested several methods for selecting the component(s) initiating the next step, and for selecting their migration. In order to evaluate our algorithms, we have tested them on various test-benches, and each test-bench was ran with various initial placements.

Based on our experiments, we can distinguish between two approaches for handling the problem. The first approach is to use algorithms with high run-time that also tend to supply good results. Due to their high run-time, these algorithms can only be ran a small number of times (with several different initial placements). The second approach is to use fast algorithms and ran them many times. We expect to get at least one good output. The first approach rely on the algorithms' ability to consider multiple local minima, thanks to their ability to escape local minimum. In the second approach the algorithms tend to stop at the first local minimum they found, however, this is compensated by the high number of runs, with many different initial placements.

Our algorithms, even for instances on which they perform poorly, achieve results not far from SA with only a fraction of its running time. This feature obviously can be very handy when one tries to get a quick estimation of the results achievable for a given instance. We believe that this work has proved the concept of selfish local search to be valid and efficient. Moreover, this concept may be useful in solving additional optimization problems arising in real-life applications.

#### REFERENCES

- [1] S. N. Adya and I. L. Markov. Combinatorial techniques for mixed-Size placement. *ACM Transactions on Design Automation of Electronic Systems*, Vol. 10(1), DOI:10.1145/1044111.1044116, 2005.
- [2] C.J. Alpert, D.P. Mehta, and S. S. Sapatnekar. *Handbook of Algorithms for Physical Design Automation*, Auerbach Publications, 2008.
- [3] Y.C. Chang, Y. W. Chang, G. M. Wu, and S. W. Wu. B\*-Trees: a new representation for non-slicing floorplans. *Proc. of the 37th Annual Design Automation Conference*, pp. 458-463, DOI: 10.1145/337292.337541, 2000.
- [4] C. Chu. *Electronic Design Automation: Synthesis, Verification, and Test*. Chapter 11: Placement. pp. 635-685. *Springer*, 2009.
- [5] E. Even-Dar and Y. Mansour. Fast Convergence of Selfish Rerouting. In *Proc. of SODA*, pp. 772-781, DOI: 10.1145/1070432.1070541, 2005.
- [6] M. Feldman, Y. Snappir, and T. Tamir. The Efficiency of Best-Response Dynamics. *The 10th International Symposium on Algorithmic Game Theory (SAGT)*, DOI: 10.1007/978-3-319-66700-3-15, 2017.
- [7] S. H. Gerez. *Algorithms for VLSI Design Automation*. *John Wiley & Sons, Inc.*, NY, USA, 1999.
- [8] P. Ghosal and T. Samanta. Thermal-aware Placement of Standard Cells and Gate Arrays: Studies and Observations, *Symposium on VLSI. IEEE Computer Society Annual*, DOI: 10.1109/ISVLSI.2008.37, 2008.
- [9] E. Huang and R. E. Korf. Optimal Rectangle Packing: An Absolute Placement Approach. *Journal of Artificial Intelligence Research*, vol. 46:47-87, 2012.
- [10] Intel Core i7 processors. [www.intel.com/content/www/us/en/products/processors/core/i7-processors.html](http://www.intel.com/content/www/us/en/products/processors/core/i7-processors.html), 2008.
- [11] Z. Jiang, H. Chen, T. Chen and Y. Chang. Challenges and Solutions in Modern VLSI Placement, *International Symposium on VLSI Design, Automation and Test (VLSI-DAT)*, DOI: 10.1109/VDAT.2007.373223, 2007.
- [12] J. Kleinberg and E. Tardos. Chapter 12: Local Search. *Algorithm Design*. Addison-Wesley, pp. 690-700, 2005.
- [13] A. Khachaturyan, S. Semenovskaya, and B. Vainshtein. Statistical-Thermodynamic Approach to Determination of Structure Amplitude Phases, *Sov. Phys. Crystallography*. Vol. 24(5):519-524, 1979
- [14] A. B. Kahng, J. Lienig, I. L. Markov, and J. Hu. VLSI Physical Design: From Graph Partitioning to Timing Closure. *Springer*, DOI: 10.1007/978-90-481-9591-6, 2011.
- [15] T. Lengauer. *Combinatorial Algorithms for Integrated Circuits*. *Wiley-Teubner*, DOI: 10.1007/978-3-322-92106-2, 1990.
- [16] Y. Lin, B. Yu, X. Xu, J. Gao, N. Viswanathan, W. Liu, Z. Li, C. J. Alpert, and D. Z. Pan. MrDP: Multiple-row Detailed Placement of Heterogeneous-sized Cells for Advanced Nodes. *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, DOI: 10.1109/TCAD.2017.2748025, 2016.
- [17] I. L. Markov, J. Hu, and M. Kim. Progress and Challenges in VLSI Placement Research. *Computer-Aided Design (ICCAD), IEEE/ACM International Conference*, DOI:10.1145/2429384.2429441, 2012.
- [18] H. Murata and E. S. Kuh. Sequence-pair based placement method for hard/soft/pre-placed modules. *Proc. of the international symposium on Physical design*, pp. 167-172, DOI:10.1145/274535.274560, 1998.
- [19] R. Norman, J. Last, and I. Haas. Solid-state Micrologic Elements, *Solid-State Circuits Conference. Digest of Technical Papers*. pp. 82-83, DOI: 10.1109/ISSCC.1960.1157264, 1960.
- [20] S. Pattanaik, S. P. Bhoi, and R. Mohanty. Simulated Annealing Based Placement Algorithms and Research Challenges. *Journal of Global Research in Computer Science*. Vol. 3(6), 2012.
- [21] N. Quinn and M. Breuer. A forced directed component placement procedure for printed circuit boards. *IEEE Transactions on Circuits and Systems*, Vol. 26(6), 1979.
- [22] S. Reda and A. Chowdhary. Effective linear programming based placement methods. *Proc. of the international symposium on Physical design*, pp. 186-191, DOI:10.1.1.83.2416, 2006.
- [23] R. A. Rutenbar. Simulated Annealing Algorithms: An Overview. *IEEE Circuits and Devices Magazine*, Vol.5(1):19 - 26, 1989.
- [24] K. Shahookar and P. Mazumder. VLSI Cell Placement Techniques. *ACM Computing Surveys*, Vol. 23(2):143-220, 1991.
- [25] Z. Yang and S. Areibi. Global Placement Techniques for VLSI Physical Design Automation. *Proc. of the 15th Intl. Conf. on Computer Applications in Industry and Engineering*, 2002.

# Integration of Polynomials over n-Dimensional Simplices

Abdenebi ROUIGUEB\*, Mohamed MAIZA†, Abderahmane TKOURT† and Imed CHERCHOUR \*

\*Ecole Militaire Polytechnique

Data Fusion and Analysis Laboratory, Algiers, Algeria

Email: rouigueb.abdenebi@gmail.com

†Ecole Militaire Polytechnique

Modeling and Optimization Techniques Laboratory, Algiers, Algeria,

Email: m\_maiza@esi.dz

**Abstract**—Integrating an arbitrary polynomial function  $f$  of degree  $D$  over a general simplex in dimension  $n$  is well-known in the state of the art to be NP-hard when  $D$  and  $n$  are allowed to vary, but it is time-polynomial when  $D$  or  $n$  are fixed. This paper presents an efficient algorithm to compute the exact value of this integral. The proposed algorithm has a time-polynomial complexity when  $D$  or  $n$  are fixed, and it requires a reasonable time when the values of  $D$  and  $n$  are less than 10 using widely available standard calculators such as desktops.

## I. INTRODUCTION

THE integral evaluation of polynomial functions over  $n$ -dimensional polytopes is essential in many applications. Particularly, it can be used to calculate the probability of a given interval of variables expressed as a polytope and when a polynomial function is used to fit the multivariate probability density function.

In dimension  $n$ , efficient integrating formulas may be set up for some types of polytopes having specific or regular shapes such as hyper-cubes and hyper-parallelepipeds. On the other hand, integrating an arbitrary polynomial function  $f$  of degree  $D$  over an arbitrary general convex polytope is a hard task.

For the simple case  $f = 1$  (when  $D = 0$ ), it turns into volume computing. Even volume computing of polytopes of varying dimension was proven to be NP-hard [2]. Hence, one can conclude that integrating of polynomials over convex polytopes is NP-hard as well, see [1] for more details.

Usually, integration over a general convex polytope is achieved by partitioning it into a finite set of simplices. Then the whole integral value can be obtained by summing the integration results of  $f$  over the resulting simplices. The computational complexity of this approach depends on the: i) triangulation algorithm complexity; ii) number of simplices; iii) integration algorithm complexity of  $f$  over a general simplex.

It has been proven that finding the smallest triangulation is NP-hard [3]. Furthermore, the number of simplices seems to increase exponentially with the dimension for all the known triangulation algorithms. Considering these challenges, it would be suitable to use a fast triangulation algorithm, and

to focus on the improvement of the integration algorithm over general simplices.

The bad news is that even integrating  $f$  over a general simplex  $\Delta$  is shown to be NP-hard [1], but the good news is that integration can be carried out within an acceptable time for some applications when  $n$  and  $D$  values are not too high ( $\leq 10$ ), and time-polynomial for some specific types of  $f$  and  $\Delta$ . For instance, this problem is polynomial time when  $n$  or  $D$  are fixed [1]. Moreover, useful efficient formulas are also given when  $f$  is quadratic and cubic and  $\Delta$  is affinely symmetric [6], when  $\Delta$  is the standard (unit) simplex and  $f$  is expressed as a product of linear forms, etc.

In [5], interesting integration formulas of arbitrary odd degree function for the  $n$ -simplex are derived using combinatorial methods. The key idea consists in employing the known integration formula over the standard simplex [8] by performing an appropriate mapping to a  $n + 1$  variable space. This method involves  $C(n + D + 1, D)$  iterations ( $C$ : combinations number). In [7], a quite similar idea by finding a suitable transformation to the standard simplex is investigated. Based on these two last studies, in this work, we want to propose a new practical algorithm for integrating a high degree (odd and even) polynomial over a general simplex, where the aim is to further accelerate the original problem transforming to another equivalent integration problem over the standard simplex.

The rest of this paper is organized as follows. The next section presents the problem statement. Then, our main contributions are described in section III. Before conclusion, complexity analysis and some experimental results are given in section IV.

## II. PROBLEM STATEMENT

Let  $\Delta \in \mathcal{R}^n$  be a general  $n$ -simplex and  $f \in Q[x_1, \dots, x_n]$  be a multivariate polynomial function of degree  $D$  with real coefficients. Commonly,  $f$  is represented as a sum of  $M$  monomial terms,  $f = \sum_{i=1}^M w_i x_1^{\alpha_1^{(i)}} \dots x_n^{\alpha_n^{(i)}}$ , where  $w_i$  and  $\alpha_j^{(i)} \in N$  ( $\alpha_1^{(i)} + \dots + \alpha_n^{(i)} \leq D$ ) correspond respectively to the coefficient and variables powers for each monomial  $i = 1..M$ .

In this study, we consider the problem of the evaluation of the multiple definite integral of  $f$  over  $\Delta$ , which we denote by  $I_{f\Delta}$ , and it is given by the formula:

$$I_{f\Delta} = \int_{\Delta} f dx_1 \dots dx_n. \quad (1)$$

We aim at providing new practical methods that compute efficiently the *exact* value of  $I_{f\Delta}$  when the coefficients of  $f$  and the vertices of  $\Delta$  are rational numbers as discussed in [1], or compute the *numerical* value when floating-point numbers are used instead. In this second case, no approximations are made and the total error in the evaluation is due only to the floating-point numbers representation precision.

### III. INTEGRATION OVER POLYTOPES

#### A. Integration over the standard simplex

The standard simplex, denoted by  $\Delta_s$ , is the polytope defined by the  $n + 1$  following vertices:  $v_0 = (0, 0, \dots, 0)'$ ,  $v_1 = (1, 0, \dots, 0)'$ ,  $v_2 = (0, 1, \dots, 0)'$ , ...,  $v_n = (0, 0, \dots, 1)'$ . The integral of  $f$  over  $\Delta_s$  is expressed as follows:

$$I_{f\Delta_s} = \int_{x_1=0}^1 \int_{x_2=0}^{1-x_1} \dots \int_{x_n=0}^{1-x_1-\dots-x_{n-1}} f dx_1 \dots dx_n. \quad (2)$$

The integral of one monomial  $x^{(\alpha)} = x_1^{\alpha_1} x_2^{\alpha_2} \dots x_n^{\alpha_n}$  can be computed efficiently using the Stroud formula [8]:

$$I_{x^{(\alpha)}\Delta_s} = \frac{\alpha_1! \dots \alpha_n!}{(n + \alpha_1 + \dots + \alpha_n)!}. \quad (3)$$

Then,  $I_{f\Delta_s}$  can be computed by summing the integral of all  $f$  monomial terms,  $I_{f\Delta_s} = \sum_{i=1}^M w_i I_{x_i^{(\alpha)}\Delta_s}$ .

Note that dynamic programming can accelerate considerably the computation. e.g. factorial terms can be computed and stored just once and used many times.

#### B. Integration over a general simplex

The majority of simplices obtained by triangulation are not standard. Integrating  $f$  over a general simplex  $\Delta$  is an NP-hard problem of varying dimension and degree [1], but it can be solved within an acceptable time for moderate dimension and degree ( $n \leq 10$ ,  $D \leq 10$ ) by the computing means which are nowadays available. To evaluate  $I_{f\Delta}$ , a good option would be to find an affine change of variables from the original space  $[x_1, \dots, x_n]$  to a new space  $[y_1, \dots, y_n]$  such that

$$I_{\Delta} = \int_{\Delta} f dx_1 \dots dx_n = \int_{\Delta_s} h dy_1 \dots dy_n, \quad (4)$$

where  $h$  is a polynomial function with the same degree as  $f$ ,  $\Delta_s$  is the standard simplex. After that, one can utilize formula 3. It is particularly noteworthy that for a non-empty volume simplex  $\Delta$ , this change of variables is always possible. We will show how to determine  $h$  terms in the rest of this section.

Let the vertices of  $\Delta$  be  $v_0 = (v_1^0, \dots, v_n^0)'$ ,  $v_1 = (v_1^1, \dots, v_n^1)'$ , ...,  $v_n = (v_1^n, \dots, v_n^n)'$ . We propose to find an affine transformation  $T : \Delta_s \rightarrow \Delta$  that maps the standard simplex  $\Delta_s$  in the  $y$  space to the general simplex  $\Delta$  in the  $x$  space, as shown in Fig 1 example. Thus,  $T$  maps each vertex

of  $\Delta_s$  to a distinct vertex of  $\Delta$ , the order of vertices is not important. Formally,  $T$  maps a given point  $\mathbf{y}$  to the point  $\mathbf{x}$  given by  $\mathbf{x} = A\mathbf{y} + B$  where  $A$  is an invertible  $n \times n$  matrix that defines the combination effect of rotation, scaling and shearing, and  $B$  is a translation vector.

The correspondence between vertices of  $\Delta_s$  and vertices of  $\Delta$  yields the following linear system  $V = A \times V_s + B$ , which is given by

$$\begin{pmatrix} v_1^0 & v_1^1 & \dots & v_1^n \\ \vdots & \vdots & \ddots & \vdots \\ v_n^0 & v_n^1 & \dots & v_n^n \end{pmatrix} = \begin{pmatrix} A_{1,1} & \dots & A_{1,n} \\ \vdots & \ddots & \vdots \\ A_{n,1} & \dots & A_{n,n} \end{pmatrix} \times \begin{pmatrix} 01 \dots 0 \\ \vdots & \ddots & \vdots \\ 00 \dots 1 \end{pmatrix} + \begin{pmatrix} B_1 \\ \vdots \\ B_n \end{pmatrix} \quad (5)$$

, where columns of matrices  $V$  and  $V_s$  represent the vertices coordinates of  $\Delta$  and  $\Delta_s$ , respectively. The unique solution of the equations system (5) is:

$$B = (v_1^0, \dots, v_n^0)' \text{ and } A_{i,j} = v_i^j - B_i \quad (\forall i, j \in 1 \dots n), \quad (6)$$

which can be computed in linear time. Affine transformations are known to preserve betweenness,  $\mathbf{x} = T(\mathbf{y})$  lies inside  $\Delta$  if and only if  $\mathbf{y}$  is inside  $\Delta_s$ . To compute  $I_{f\Delta}$ , we propose to perform the change of variables  $\mathbf{x} = A \times \mathbf{y} + B$  where the  $i^{th}$  component of  $\mathbf{x}$  is

$$\begin{aligned} x_i &= A_{i,1} \times y_1 + \dots + A_{i,n} \times y_n + B_i \\ &= \sum_{j=1}^n A_{i,j} \times y_j + B_i. \end{aligned} \quad (7)$$

Hence, we have:

$$\begin{aligned} I_{f\Delta} &= \int_{\Delta} f(x_1, \dots, x_n) dx_1 \dots dx_n \\ &= \int_{\Delta_s} f(\sum_{j=1}^n A_{1,j} \times y_j + B_1, \dots, \sum_{j=1}^n A_{n,j} \times y_j + B_n) \\ &\quad |det(Jac_{T(y_1, \dots, y_n)})| dy_1 \dots dy_n, \end{aligned} \quad (8)$$

where  $|det(Jac_{T(y_1, \dots, y_n)})| = |det(A)|$  is the absolute value of the determinant of the Jacobian of  $T$ .

For the simple case of a single monomial term,  $f = x^{(\alpha)} = x_1^{\alpha_1} \dots x_n^{\alpha_n}$  we have:

$$\begin{aligned} I_{x^{(\alpha)}\Delta} &= \int_{\Delta_u} (\sum_{j=1}^n A_{1,j} y_j + B_1)^{\alpha_1} \dots (\sum_{j=1}^n A_{n,j} y_j + B_n)^{\alpha_n} \\ &\quad |det(A)| dy_1 \dots dy_n. \end{aligned} \quad (9)$$

Let  $P_{x_i} = \sum_{j=1}^n A_{i,j} y_j + B_i$  be the dense polynomial of degree one in the  $n$ -dimensional space of  $y$  variables. The desired polynomial function  $h$  is then equal to  $|det(A)| f(P_{x_1}, \dots, P_{x_n})$ .

Therefore, determining  $h$  coefficients can be carried out according to the expression of  $f$  by performing a series of additions and multiplications of dense intermediate polynomials of degree  $d \leq D$  ( $D = \text{degree of } f$ ).

For example, as illustrated in Fig. 1, to integrate  $f = x_1 x_2^3 + x_1^2 x_2 + x_2^2 + 2x_1 x_2 + x_1 + 2$  over the simplex  $\Delta$  defined by

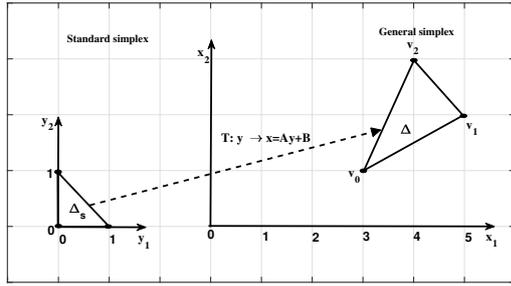


Fig. 1. 2-Dimensional affine transformation example.

the vertices  $V = \{v_0 = (3, 1)', v_1 = (5, 2)', v_2 = (4, 3)'\}$ , by using equations (6,7), we obtain:

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, B = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \begin{matrix} p_{x_1} = 2y_1 + y_2 + 3 \\ p_{x_2} = y_1 + 2y_2 + 1 \end{matrix}$$

Determining  $h$  depends indeed on the regular expression used to denote  $f$ ; in other words, it depends on the position of brackets and operators '\*', '+' in  $f$  expression. For this example,  $h$  can be built efficiently according to the scheme shown in Fig. 2:(a) when  $f$  is expressed as a sum of monomials, or according to the scheme of Fig. 2:(b) when  $f$  is factorized as follows  $f = (x_1x_2 + 1) * (x_2^2 + x_1 + 2)$ . As you can see in Fig. 2:(a), terms  $x_1x_2$  and  $x_2^2$  are computed just once, and used to compute the four first terms of  $f : x_1x_2^3, x_1^2x_2, x_2^2, 2x_1x_2$

For this approach, integration efficiency depends conjointly on : i) the factorization tree of  $f$  and ii) the complexity of addition and multiplication of intermediate polynomials. Finding the optimal factorization tree for an arbitrary dense polynomial is a difficult combinatorial problem because the search space of all possible factorization schemes is large; e.g. using straight-line-programs [1] may be useful. In this study, our main contribution is focused on the second point; on the proposition of new methods to accelerate furthermore addition and particularly multiplication of intermediate polynomials involved during the construction of  $h$ .

### C. Discussion

In the proposed method, computing  $I_{f\Delta}$  is achieved by performing a suitable change of variables from  $x$  to  $y$  space such that equation (4) holds. And then, integrating the monomial terms of  $h$  over the standard simplex  $\Delta_s$  using formula (3) in a polynomial time.  $h$  terms are determined by accomplishing a long sequence of addition and multiplication of *dense* intermediate polynomials of degree up to  $D$ . Almost all obtained polynomials are dense because the matrix transformation  $A$  is often not sparse for general simplices resulting from the triangulation process. The big part of the computational complexity is due to multiplication rather than addition of intermediate polynomials. Indeed, the multiplication  $P = P_1 \times P_2$  of two polynomials represented as a sum of monomial terms can be carried out in two steps: i) distribution and monomial multiplication (a Cartesian product) and ii) simplification of

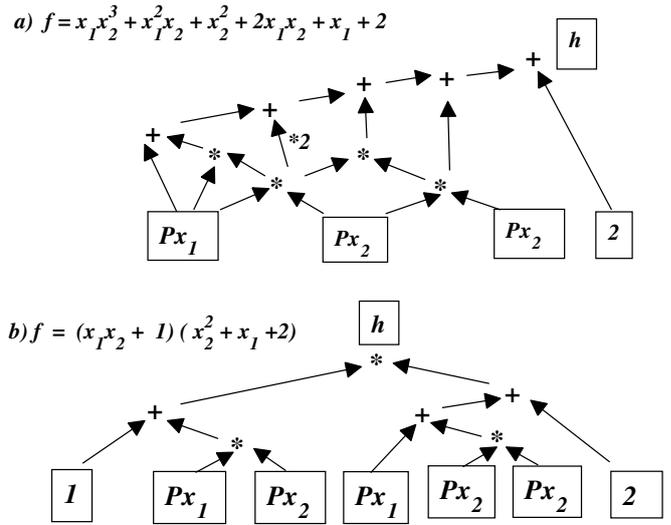


Fig. 2. Example of variables change according to 2 different factorizations.

result terms having the same degree. It should be emphasized that the simplification (step ii) requires more of computational complexity than the distribution and multiplication (step i). Let  $k_1, k_2$  be sizes of  $P_1, P_2$ , respectively. Consequently, we need  $k_1 \times k_2$  elementary operations of multiplication in step i), but we need  $k_1 \times k_2 \times C$  operations where  $C$  is the cost of the simplification of a given monomial term  $m$ .  $C$  is equal to the cost of search plus the cost of insertion of term  $m$  within the structure of  $P$ .

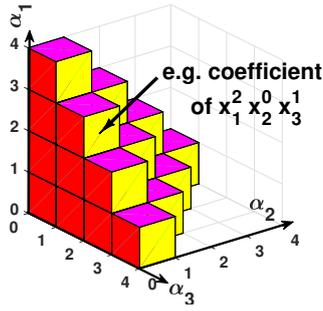
Therefore, it is interesting to improve monomial simplification by the proposition of suitable polynomial representation and efficient algorithms for multiplication.

### D. Proposition 1: accelerating variables change

The physical memory can in fact be seen as a continuous sequence of cells and all intermediate polynomials terms are stored of course in that space. During the variables change process, we need to find efficiently many times the location of a monomial given their variable powers, hence the order of monomials in memory is important.

As all the polynomials obtained are dense, we propose to represent a given polynomial  $P$  as a triplet  $(n, d, W)$  where  $n$  is the dimension,  $d$  is its degree ( $d \leq D$ ), and  $W$  is a vector of floats containing the monomial coefficients of  $P$ . The size of  $W$  is then equal to  $(n + d)! / (n!d!)$ . It is a good idea to save only the monomial coefficients into a compact structure according to a *particular order* and to not save variables powers ( $\alpha$  vectors). The desired order must allow a fast mapping in both directions between variables powers vector and the monomial position in the structure of  $P$ . Consequently, space complexity will be reduced because variables powers ( $\alpha$  vectors) are not saved.

The memory position (index) of the coefficient of a given monomial  $x^{(\alpha)} = x_1^{\alpha_1} \dots x_n^{\alpha_n}$  must be calculated efficiently based on the values of  $n, d, \alpha$  and the chosen order. To this end, we consider that  $w x_1^{\alpha_1} \dots x_n^{\alpha_n}$  is ordered before (on the

Fig. 3. VOIS structure example ( $D=3, n=3$ ).**Procedure 1** Mapping powers to index ( $Pow2Ind$ )

---

**Input:**  $\alpha$  monomial powers vector  
degree  $d$  and dimension  $n$  of the polynomial

**Output:**  $Ind$  (the corresponding index memory of  $x^{(\alpha)}$ )  
 $Ind \leftarrow 0$ ;

- 2:  $j \leftarrow n - 1$ ;  $i \leftarrow d$ ; start cell in Pascal square  
 $k \leftarrow 0$ ; variable index in vector  $\alpha$
- 4: **while**  $\alpha \neq (0, \dots, 0)$  **do**
- 6:   **if**  $\alpha[k] > 0$  **then**  
       $Ind \leftarrow Ind + Pascal[j, i]$ ; increment  $Ind$   
       $i \leftarrow i - 1$ ; shift left in Pascal  
       $\alpha[k] \leftarrow \alpha[k] - 1$ ;
- 8:   **else**  
       $j \leftarrow j - 1$ ; shift up in Pascal  
       $k \leftarrow k + 1$ ; shift right in  $\alpha$
- 10:   **end if**

**end while**

---

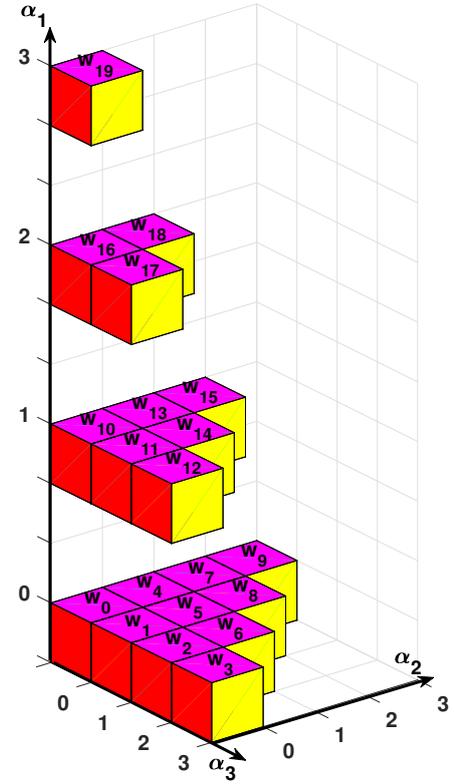
left) of  $w'x_1^{\alpha'_1} \dots x_n^{\alpha'_n}$  if  $[\alpha_1 \dots \alpha_n] < [\alpha'_1 \dots \alpha'_n]$ , which can be evaluated recursively as follows:

$$[\alpha_1 \dots \alpha_n] < [\alpha'_1 \dots \alpha'_n] \text{ if } \begin{cases} \alpha_i < \alpha'_i \text{ or } (\alpha_i = \alpha'_i \text{ and} \\ [\alpha_{i+1} \dots \alpha_n] < [\alpha'_{i+1} \dots \alpha'_n]) \end{cases}$$

We propose to represent the coefficients  $W$  of a dense polynomial of degree  $d$  and  $n$  variables as a Virtual Ordering Integer Simplex which we refer to as VOIS structure in this paper. The coefficient of the monomial  $x_1^{\alpha_1} \dots x_n^{\alpha_n}$  is stored in the cell  $(\alpha_1, \dots, \alpha_n)$  of the VOIS structure. Fig 3 illustrates an example for  $d=3$  and  $n=3$ . The coefficients of  $P$  are stored in the physical memory as a flat vector according to the VOIS order as follows:  $[w_0, w_1, \dots, w_k]$  where  $k = (n + d)! / n! * d!$  is the size of  $P$ , as illustrated in Fig 4.

For polynomial multiplication, we need both to handle monomials powers  $(\alpha_1, \dots, \alpha_n)$ , and to access directly to their memory location. The principal aim of the VOIS structure is to speed up conversion between monomial powers vectors and theirs corresponding memory indices.

The proposed functions that map powers vector to the corresponding memory index ( $pow2ind$ ) and the inverse mapping ( $ind2pow$ ) are displayed in Algorithms 1 and 2, respectively.

Fig. 4. Memory indices of monomials coefficients ( example:  $D=3, n=3$ ).**Procedure 2** Mapping index to powers ( $Ind2Pow$ )

---

**Input:**  $Ind$  monomial powers vector  
degree  $d$  and dimension  $n$  of the polynomial

**Output:**  $\alpha$  monomial powers vector  
 $\alpha \leftarrow [0, \dots, 0]$ ;

- 2:  $j \leftarrow n - 1$ ;  $i \leftarrow d$ ;
- 4: **while**  $Ind \neq 0$  **do**
- 6:   **if**  $Ind \geq Pascal[j, i]$  **then**  
       $\alpha[k] \leftarrow \alpha[k] + 1$ ;
- 8:   **else**  
       $j \leftarrow j - 1$ ; shift up in Pascal  
       $k \leftarrow k + 1$ ; shift right in  $\alpha$
- 10:   **end if**

**end while**

---



TABLE I  
DEFINING CHARACTERISTICS OF FIVE EARLY DIGITAL COMPUTERS

#	Method	Search time	Insertion time	Space complexity
1	Sorted linked list	$O(X)$ //check all elements	$O(1)$	$X^{*(n+1)}$ //Coefficient and powers ( $\alpha$ )
2	Sorted dynamic array	$O(\text{Log}(X))$ //dichotomic search	$O(X)$ //shift elements right	$X^{*(n+1)}$ //Coefficient and powers ( $\alpha$ )
3	Static pre-allocated array	$O(\text{Log}(X))$ //dichotomic search	$O(1)$	$X^{*(n+1)}$ //Coefficient and powers ( $\alpha$ )
4	<b>VOIS without indices saving</b>	<b><math>O(n+d)</math></b> //Pow2Ind	<b><math>O(1)</math></b>	<b><math>X</math></b> //only coefficients
5	<b>VOIS with indices saving</b>	<b><math>O(1)</math></b> //direct access	<b><math>O(1)</math></b>	<b><math>X + [((n+1)d)/(d+n)]*X</math></b> //coefficients + indices trace

where  $X = (n + d)!/(n!d!)$  is equal to the number of all polynomial terms of  $P(n, d)$ .

**Procedure 4** Polynomial multiplication with indices saving(*PolMul*)

**Input:**  $P_1, P_2$  two polynomials with degrees  $d_1, d_2 = 1$ , in dim.  $n$

**Output:**  $P = P_1 \times P_2$

$P \leftarrow \text{ZerosPolynomial}; m \leftarrow 0;$

2: **if**  $Ix_{d_1}$  is empty **then**

4: **for**  $i_1 = 1$  **to** Size of  $P_1$  **do**

$\alpha_1 \leftarrow \text{Ind2Pow}(i_1, d_1, n)$

6: **for**  $i_2 = 1$  **to** Size of  $P_2$  **do**

$\alpha_2 \leftarrow \text{Ind2Pow}(i_2, d_2, n); \alpha \leftarrow \alpha_1 + \alpha_2;$

$i \leftarrow \text{Pow2Ind}(\alpha, d, n)$

$P[i] \leftarrow P[i] + P_1[i_1] \times P_2[i_2];$

$Ix_{d_1}[m] \leftarrow i; m \leftarrow m + 1; //$  save  $i$  sequence

8: **end for**

**end for**

10: **else**

12: **for**  $i_1 = 1$  **to** Size of  $P_1$  **do**

14: **for**  $i_2 = 1$  **to** Size of  $P_2$  **do**

$i \leftarrow Ix_{d_1}[m]; m \leftarrow m + 1; //$ direct access

$P[i] \leftarrow P[i] + P_1[i_1] \times P_2[i_2];$

16: **end for**

**end for**

18: **end if**

$$\leq (D + n)!/(D!n!) * (n + 1) * O(n + D)$$

$$\leq (D + n)^D/(D!) * (n + 1) * O(n + D)$$

$$\leq O(n^D) * O(n) * O(n + D)$$

$$\leq O(n^{D+2})$$

, which is polynomial with a varying  $n$ . The complexity of polynomial addition is also polynomial in time; it is lesser than the multiplication complexity. According to the used polynomial factorization, we need at most to compute  $(D+n)!/(D!n!)$  multiplications and  $(D+n)!/(D!n!)$  additions which is polynomial for a fixed  $D$ . The total complexity of integrating is then polynomial because the composition of two polynomial functions is also polynomial.

The permutation between  $D$  and  $n$  in the given proof allows us to conclude that the complexity is polynomial for a fixed number of variables. This result agrees with results given recently in [1].

Table II shows the measured integration time using the proposed VOIS method for some examples of polynomials and simplices generated randomly when varying  $D$  and  $n$ . Experiment are carried out on a standard computer. The best integration results for less than a second are highlighted in bold and the worst results exceeding 10 hours are not displayed. One can notice that time-complexity increases very fast as an exponential when augmenting together  $D$  and  $n$ , but it increases with a lower rate when either  $D$  or  $n$  are low. We also notice that saving indices allows reducing time by a factor nearby 5.

## V. CONCLUSION

In this paper, we have proposed an integration method of high dimensional polynomial functions with high degree over a general simplex by performing an affine change of variables to the standard simplex where efficient formulas are already known.

The suggested variables change turns into making addition and multiplication operations over intermediate polynomials many times. To this end, we have proposed a compact data structure for polynomial representation, that we have called VOIS, in order to optimize the different operations involved during the polynomial transforming from the general simplex integration problem to an equivalent standard simplex integration problem.

For a fixed degree (and a varying dimension) or for a fixed dimension (and a varying degree) the integration computational complexity of our algorithm over a general simplex

output polynomial  $P$  of multiplication. One can see that our proposition (methods 4 and 5) overcomes incontestably the three first classical methods in terms of the trade-off between time and space complexity.

Another finding was that the global complexity of our algorithm for the integration over a general simplex is polynomial in time when the degree  $D$  or the dimension  $n$  are fixed. The proof of this result is as follows. In our proposition, all polynomial multiplications are carried out between a polynomial  $p_1$  ( $d_1 \leq D$ ) with a one degree polynomial  $P_2$  ( $d_2 = 1$ ).

For a fixed degree  $D$ , the complexity of multiplication  $P_1 \times P_2$  is:  $\text{CompMul} = \text{size of } P_1 * \text{size of } P_2 * O(n+d)$ , where  $O(n + d)$  is the search-insertion time complexity of a one monomial term (see row 4 in Table I). Hence, we have:

$$\text{CompMul} = (d_1 + n)!/(d_1!n!) * (n + 1) * O(n + d)$$

TABLE II  
TIME OF INTEGRATION OF A DENSE POLYNOMIAL OVER A GENERAL SIMPLEX

Degree	Method1: VOIS with indices saving (sec)								Method2: VOIS without indices saving (sec)							
	$n = 2$	$n = 4$	$n = 6$	$n = 7$	$n = 8$	$n = 9$	$n = 12$	$n = 15$	$n = 2$	$n = 4$	$n = 6$	$n = 7$	$n = 8$	$n = 9$	$n = 12$	$n = 15$
$D = 2$	<b>0</b>	<b>0</b>	<b>0.001</b>	<b>0.004</b>	<b>0.027</b>	<b>0.227</b>	287.81	-	<b>0</b>	<b>0</b>	<b>0.001</b>	<b>0.004</b>	<b>0.026</b>	<b>0.225</b>	291.71	-
$D = 4$	<b>0</b>	<b>0.001</b>	<b>0.005</b>	<b>0.015</b>	<b>0.054</b>	<b>0.258</b>	297.77	-	<b>0</b>	<b>0.015</b>	<b>0.027</b>	<b>0.045</b>	<b>0.126</b>	<b>0.447</b>	293.87	-
$D = 6$	<b>0</b>	<b>0.005</b>	<b>0.095</b>	<b>0.348</b>	1.061	3.245	333.65	-	<b>0</b>	<b>0.013</b>	<b>0.348</b>	1.339	4.509	13.64	520.79	-
$D = 7$	<b>0</b>	<b>0.011</b>	<b>0.340</b>	1.334	4.446	16.25	666.91	-	<b>0</b>	<b>0.033</b>	1.278	5.648	21.69	74.25	2133.6	-
$D = 8$	<b>0</b>	<b>0.023</b>	<b>0.966</b>	4.607	19.19	74.17	2889.1	-	<b>0</b>	<b>0.080</b>	4.199	4.199	93.76	363.2	-	-
$D = 9$	<b>0</b>	<b>0.050</b>	2.67	14.66	71.50	324.45	-	-	<b>0.001</b>	<b>0.178</b>	12.23	72.281	361.4	1584.1	-	-
$D = 12$	<b>0.001</b>	<b>0.318</b>	39.12	322.59	10534	-	-	-	<b>0.002</b>	1.248	185.2	1534.5	10835	-	-	-
$D = 15$	<b>0.002</b>	1.441	365.66	18928	-	-	-	-	<b>0.004</b>	6.135	1727.3	19647	-	-	-	-
$D = 20$	<b>0.005</b>	11.25	34579	-	-	-	-	-	<b>0.012</b>	50.09	34415	-	-	-	-	-

In experiments, only one core of the Processor Intel i7 3.7 GHz is used.

is polynomial. However, when varying both dimension and degree, the complexity in experiments that we have carried out seems to increase exponentially. In this last case, we recall that integration of a general polynomial function over a general simplex is shown to be NP hard [1].

A second aspect not fully examined in this work relates to the representation of the input polynomial; more precisely to the regular expression used to represent the polynomial that we want to integrate. For instance, we have found in some experiments that representing the polynomial function as a product of lower-degree polynomials, if it is possible, is more efficient than using a dense form expressed as a sum of monomial terms. We recommend orienting future works on the problem of determining the optimal factorization tree of polynomial functions in relation to integration performances.

REFERENCES

[1] V. Baldoni and N. Berline and J. A. De Loera and M. Köppe and M. Vergne, "How to Integrate a Polynomial over a Simplex," *Math. Comput. J.*, vol. 80, 2011, pp. 297–325.  
 [2] M. E. Dyer and A. M. Frieze, "Frieze, On the complexity of computing the volume of a polyhedron," *SIAM J. Comput.* vol. 17, no. 5, 1961, pp. 967-974.  
 [3] J. A. De Loera, J. Rambau, and F. Santos, *Triangulations: Structures and algorithms*, Book manuscript, 2008.  
 [4] A. H. Stroud, *Approximate Calculation of Multiple Integrals*. Prentice-Hall, Englewood Cliffs, NJ, 1971.  
 [5] A. Grundmann and H. M. Moller, "Invariant Integration Formulas for the n-Simplex by Combinatorial Methods," *SIAM J. Numer. Anal.* vol. 17, no. 5, 1961, pp. 282-290.  
 [6] P. C. Hammer and A. H. Stroud, "Numerical integration over simplexes," *Math Tables other Aids Comput.* vol. 10, 1956, pp. 137-139.  
 [7] F. Bernardini, "Integration of polynomials over n-dimensional polyhedra," *Computer-Aided Design* vol. 23, no. 11, 1991, pp. 51-58.  
 [8] A. H. Stroud, *Approximate Calculation of Multiple Integrals*, Prentice-Hall, Englewood Cliffs, NJ, 1971.  
 a product of linear forms



# Customized Genetic Algorithm for Facility Allocation using $p$ -median

S. D. de S. Silva  
Universidade Federal do Amazonas  
Manaus, Brasil  
Email: srgio.deo@gmail.com

M. G. F. Costa  
Universidade Federal do Amazonas  
Manaus, Brasil  
Email: mcosta@ufam.edu.br

C. F. F. Costa Filho  
Universidade Federal do Amazonas  
Manaus, Brasil  
Email: ccosta@ufam.edu.br

□ **Abstract**—The  $p$ -median problem is classified as a NP-hard problem, which demands a long time for solution. To increase the use of the method in public management, commercial, military and industrial applications, several heuristic methods has been proposed in literature. In this work, we propose a customized Genetic Algorithm for solving the  $p$ -median problem, and we present its evaluation using benchmark problems of OR-library. The customized method combines parameters used in previous studies and introduces the evolution of solutions in stationary mode for solving PMP problems. The proposed Genetic Algorithm found the optimum solution in 37 of 40 instances of  $p$ -median problem. The mean deviation from the optimal solution was 0.002% and the mean processing time using CPU core i7 was 17.7s.

## I. INTRODUCTION

Facility location problems (FLP) are usually employed for solving public, commercial, industrial and military problems. In these problems, service demand points must be attended by a limited number of facilities. The  $p$ -median problem (PMP) is a type of FLP problem that aims searching a given location that minimizes the sum of the distances between  $N$  demand points and the nearest facility [1].

The computational complexity theory classifies the PMP as a non-polynomial hard problem (NP-hard problem). Meta-heuristic methods are usually used for solving NP-hard problems whose optimal solution method does not exist or is not known: Greedy Interchange (GI) [2], Neighborhood (N) and Exchange [2], Semi-Lagrangian relaxation [3], Simulated Annealing (SA) [4], Tabu Search (TS) [5], Genetic Algorithm (GA) [6-7].

To enable comparative studies of these methods, benchmarking data bases are used. The Operational Research (OR) library [8] and the Traveling Salesman Problem (TSP) [9] are the most used ones [6,7,10,11,12,13].

In the comparisons made in [6] and [11], the GA heuristic stands out as the best one in terms of time and precision of solution. Nevertheless, concerning the precision of the solutions, the GAs presented in these works have a worse result than the customized GAs, presented in [7], as well as when compared to GA combinations with other heuristics, presented in [11].

In [6], the authors used the OR-library [14] and two others more simple databases to evaluate several methods used in PMP solution: ADE (Alp, Drezner and Erkut) GA, Gamma Heuristic (GH), SA, Myopic, Exchange and N. The algorithm known as ADE GA performs a greedy search using the genetic material of two individuals randomly selected, evaluating all the possible combinations of generated offspring. The algorithm found solutions with an average distance from the optimal solution (OPT solution) of 0.41%, in 85% of the OR-library problems, and an average time of 18 seconds.

In [11] the authors performed a comparative study of a GA, an N algorithm and a hybrid GA and N algorithm, using the TSP-library. The GA proposed by the authors is similar to ADE GA, differing only in the use of an algebraic method to select a pair of parents. The GA converged to a solution in less time than the other heuristics. The CPU average time was 126.8min. The GA presented solutions with an average distance from the OPT solution of 0.000016%, and an average time of 391.5min.

In [7], a simple GA was compared to ADE GA, using a subset of OR-library. This GA investigates the use of  $p$  centroids to find the initial solutions of the algorithm. The GA found OPT solutions in 14 of the 15 subset problems. The average CPU time was 60.1s and 0.2s, for the simple GA and ADE GA, respectively. The average deviation was 0.007% and 0.02% for simple GA and ADE GA, respectively.

Table I shows a summary of the main characteristics of the GA used for solving the  $p$ -median problem in [6, 7, 11].

This work aims investigating the customization of GA for solving PMP problems. Three steps of the GA are investigated: selection operator, crossover operator and population updating. This investigation has the objective of generating a best performance of GA in finding OPT solutions for PMP problems.

The random selection operator employed in [6] and [11] does not take into account the individual's fitness when they are selected for crossover. The ranking selection operator employed in [7] assigns a selection probability to individuals directly proportional to their position in a ranking of the fitness function. In this work we investigate the use of the roulette wheel selection operator. The difference between the

□ This work was supported by *Samsung Eletronica da Amazonia*, under the terms of the Brazilian Federal Law number 8.387/91.

TABLE I.  
GENETIC ALGORITHM CHARACTERISTICS USED FOR SOLVING THE P-MEDIAN PROBLEM IN [6, 7, 11].

Paper	Data base	GA characteristics			Results	
		Selection	Crossover	Heuristics studied	GA Deviation from OPT	Faster Heuristic
[6]	OR - Library, Alberta, Galvão e Koerkel	Random	Merging	ADE, GH, SA, Myopic, Exchange, BV	Up to 0.41% from OPT at 85% of OR problems. 0% at Alberta problems.	ADE
[11]	TSP – Library	Random	Merging	GA [11], BV, Hybrid between GA [11] and N	Up to 0.008% from OPT at 100% of TSP problems.	GA [11]
[7]	OR – Library (15 problems)	Ranking-based	Partial Match	ADE, GA [7]	GA [7]	ADE GA

ranking operator and the roulette wheel selection operator is that, the last one assign individuals a selection probability directly proportional to their fitness value.

We also propose using the single-point crossover operator. Differently from the merging operators [6,11] and partial match operators [7], the single-point crossover operator generate offspring without evaluating the parents. This implies in less processing time demand.

At last, we propose use a steady-state population updating [15]. In this updating mode, the fitness of children is compared to their parent's fitness. When the fitness value of the offspring is lower than their father's fitness, they are discarded. Offspring with better fitness values than their fathers are preserved with a probability of 75%.

The results obtained in this study are compared with the results obtained with the ADE GA [6] and simple GA [7]. For this comparison, we employed PMPs of OR-library and did a benchmark of the machines used for simulations in these previous works.

## II. METHODS

### A. Proposed Genetic Algorithm

Genetic Algorithm is a stochastic optimization algorithm, inspired by the theory of evolution of Charles Darwin [16]. Since its proposition, it has been effectively applied in the solution of complex problems, like TSP [9] and PMP [6,7,11].

In GA, initially, a population of chromosomes is randomly generated. In the sequence, the individuals of this population are modified by applying evolution operators, iteratively. A chromosome represents a solution to the problem. The fitness value of each chromosome is evaluated through an objective function of the problem.

The implementation of GA usually consists of three steps: the definition of the genetic codification model, the definition of the objective function and the parameterization of the evolution operators.

In this work, the genetic codification model uses the facility indexes and the objective function is given by the PMP. The structure of the proposed GA is presented in the steps of Algorithm 1.

### Algorithm 1 Proposed Genetic Algorithm

*Begin*

*Randomly generate the initial population*

*Compute fitness of population*

*Repeat for x generations*

*Roulette wheel selection of 2 parents*

*One-point crossover, at a 95% probability*

*One-gene random mutation, at a 5% probability*

*Compute fitness*

*Replace the parents with lower fitness than the children, at a 75% probability*

*Introduce a random chromosome to the population*

*Until population has converged*

*End*

### Genetic codification

As stated before, the genetic codification uses the facility indexes. The same approach was also used in [6, 7, 11]. Figure 1 shows an encoded chromosome representing a solution in a PMP problem with 8 facilities to be allocated among 100 possible locations.

1	20	31	4	76	91	62	100
---	----	----	---	----	----	----	-----

Fig. 1 Example of an encoded chromosome used in a PMP problem with 8 facilities

### Compute fitness

According to equation 1, the goal of the PMP is minimize  $f$ : the sum of the distances between the demand points and the nearest facility.

$$f = \sum_{i=1}^n \sum_{j=1}^p d_{ij} x_{ij} \quad (1)$$

$d_{ij}$  = distance between point  $i$  and point  $j$

$$x_{ij} = \begin{cases} 1, & \text{if demand } i \text{ is attended by } j \\ 0, & \text{otherwise} \end{cases}$$

$n$  = number of total locations

$p$  = number of medians

*Roulette Wheel Selection*

The selection operator used is the roulette wheel operator [17,18]. In this study, the selection operator assigns a probability value to each individual that is inversely proportional to its fitness value. The inversely dependence is due to the fact that, in the  $p$ -median problem, best individuals are those with lower  $f$  values, given in equation (1). Therefore, fitter individuals are the most likely to have children. This behavior favors the generation of more fit individuals. Table II illustrates the probability values used by different selection operators for four individuals.

TABLE II.  
SELECTION OPERATORS CHARACTERISTICS

Chromosome	Fitness Value	Probability of selection operator (%)		
		Roulette Wheel	Random	Ranking
1	200	28.8	25	30
2	900	6.3	25	20
3	100	57.6	25	40
4	800	7.2	25	10

*One-point Crossover*

The one-point crossover operator is used in this study [18]. This operator randomly generates a reference point to permute genes between fathers. The crossover probability used is 95%. Figure 2 illustrates the genetic permutation performed by the one-point crossover operator. To avoid repeated indexes in the offspring, we do a scan in the genes of each child, and replace the repeated index with another value randomly selected.

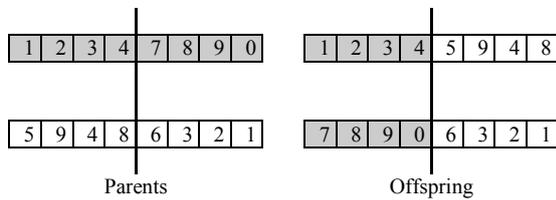


Fig. 2 Illustration of one-point crossover operation

*One-gene random mutation*

The mutation operator used randomly selects one gene [7], with probability of 5%, and performs a mutation. Figure 3 illustrates the mutation operator. One gene with index value of 5 is selected and replaced with the index value of 7. The replacing value is random selected.

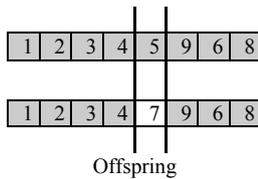


Fig. 3 Illustration of one-gene mutation operation

*B. Metrics*

In this study, the metrics used for performance evaluation of GA algorithms are the number of optimal solutions found, the percentual deviation of a non-optimal solution from the OPT solution, and the algorithm convergence time.

*C. PMP data base for benchmarking*

Aiming to compare the results obtained in this study with the results obtained in other two works [6,7], we use the PMP section of OR-library. For each problem, are given: the number of points,  $N$ ; the number of facilities,  $p$ ; the OPT solution and a matrix with the distances between each pair of points.

*D. CPU benchmarking*

The algorithm convergence time depends on the CPU model and the clock of the machine used for its implementation. Therefore, to compare the results of the GA used in this study with the GAs used in [6,7], we performed the benchmark between CPUs using the Dhrystone (D) method [19].

Using a default algorithm with integer numbers, the D method assigns a numeric value to each CPU. This value represents the number of millions of Dhrystone instructions processed per second (DMIPS) per MHz of clock. The DMIPS value of the machine used in this study (Core I7 7500U) is made equal to 1. The DMIPS values of the machines used in [6,7] are then divided by it and ratio values are obtained. The last column of Table III shows these ratio values. As shown, the machines used in [6] and [7] process 15.1% and 104.1%, respectively, of the DMIPS processed by the machine used in this study.

TABLE III.  
BENCHMARK COMPARISON OF THREE DIGITAL COMPUTERS

CPU	Clock (MHz)	DMIPS/ MHz	Product	Ratio
Pentium III [6]	733	3.4	2492.2	0.151
Core I7 4770K [7]	2000	8.57	17140	1.041
Core I7 7500U	1800	9.1	16380	1
This study				
Product = (Clock*DMIPS/MHz)		Ratio=(Product/16380)		

III. RESULTS

Table IV shows, for the GA proposed in this study, and for the GAs proposed in [6] and [7], the following results: the number of OPT solutions; the percentual deviations from non-optimal solutions to OPT solutions and the GA algorithm processing time. The processing time of the GAs proposed in [6] and [7] are multiplied by the ratio value shown in Table III. Similarly to [6], the results of this study were produced by a C++ code, implementing 10 runs for each one of the 40 OR-library PMP problems. The best results are in bold.

TABLE IV.  
EXPERIMENTAL RESULTS

Problem	N	p	Optimal Solution	Number of solutions	p/N (%)	ADE GA [6]		GA [7]		GA proposed (GAP)		Best deviation from optimal solution		
						Fitness value	Time (s)	Fitness value	Time (s)	Fitness value	Time (s)	ADE GA	GA	GAP
Pmed1	100	5	5819	75287520	5.00	OPT*	0.015	OPT	0.104	OPT	<b>0.001</b>	0	0	0
Pmed2	100	10	4093	1.73E+13	10.00	OPT	0.015	OPT	0.94	OPT	<b>0.008</b>	0	0	0
Pmed3	100	10	4250	1.73E+13	10.00	OPT	0.03	OPT	0.209	OPT	<b>0.003</b>	0	0	0
Pmed4	100	20	3034	5.36E+20	20.00	OPT	0.03	OPT	1.3	OPT	<b>0.026</b>	0	0	0
Pmed5	100	33	1355	2.95E+26	33.00	OPT	<b>0.045</b>	OPT	3.3	OPT	0.046	0	0	0
Pmed6	200	5	7824	2.54E+09	2.50	OPT	0.06	OPT	2.7	OPT	<b>0.005</b>	0	0	0
Pmed7	200	10	5631	2.25E+16	5.00	OPT	0.075	OPT	4.1	OPT	<b>0.026</b>	0	0	0
Pmed8	200	20	4445	1.61E+27	10.00	OPT	<b>0.105</b>	OPT	14.8	OPT	0.129	0	0	0
Pmed9	200	40	2734	2.05E+42	20.00	OPT	<b>0.181</b>	OPT	32.3	OPT	0.519	0	0	0
Pmed10	200	67	1255	1.45E+54	33.50	1256	0.301	OPT	41.4	OPT	1.2	0.080	0	0
Pmed11	300	5	7696	1.96E+10	1.67	OPT	0.256	OPT	28.8	OPT	<b>0.002</b>	0	0	0
Pmed12	300	10	6634	1.4E+18	3.33	OPT	0.181	OPT	47.8	OPT	<b>0.066</b>	0	0	0
Pmed13	300	30	4374	1.73E+41	10.00	OPT	<b>0.316</b>	OPT	78.4	OPT	0.64	0	0	0
Pmed14	300	60	2968	9.04E+63	20.00	OPT	<b>0.663</b>	OPT	301.8	OPT	2.9	0	0	0
Pmed15	300	100	1729	4.16E+81	33.33	1733	0.949	1731	343.6	OPT	14.8	0.231	0.116	0
Pmed16	400	5	8162	8.32E+10	1.25	OPT	0.346	-	-	OPT	<b>0.009</b>	0	-	0
Pmed17	400	10	6999	2.58E+19	2.50	OPT	0.361	-	-	OPT	<b>0.096</b>	0	-	0
Pmed18	400	40	4809	1.97E+55	10.00	OPT	<b>0.843</b>	-	-	OPT	0.999	0	-	0
Pmed19	400	80	2845	4.23E+85	20.00	2846	2	-	-	OPT	42.2	0.035	-	0
Pmed20	400	133	1789	1.3E+109	33.25	1792	0.949	-	-	OPT	15.95	0.168	-	0
Pmed21	500	5	9138	2.55E+11	1.00	OPT	0.572	-	-	OPT	<b>0.016</b>	0	-	0
Pmed22	500	10	8579	2.46E+20	2.00	OPT	0.678	-	-	OPT	<b>0.107</b>	0	-	0
Pmed23	500	50	4619	2.31E+69	10.00	OPT	2.4	-	-	OPT	<b>2.31</b>	0	-	0
Pmed24	500	100	2961	2E+107	20.00	2962	3.2	-	-	OPT	15.7	0.034	-	0
Pmed25	500	167	1828	7.9E+136	33.40	1832	4.8	-	-	OPT	105.9	0.219	-	0
Pmed26	600	5	9917	6.37E+11	0.83	OPT	1	-	-	OPT	<b>0.013</b>	0	-	0
Pmed27	600	10	8307	1.55E+21	1.67	OPT	1.2	-	-	OPT	<b>0.16</b>	0	-	0
Pmed28	600	60	4498	2.77E+83	10.00	4499	3.7	-	-	OPT	23.96	0.022	-	0
Pmed29	600	120	3033	1E+129	20.00	3035	6.6	-	-	OPT	93.422	0.066	-	0
Pmed30	600	200	1989	2.5E+164	33.33	1997	11.9	-	-	OPT	251.54	0.402	-	0
Pmed31	700	5	10086	1.38E+12	0.71	OPT	2.2	-	-	OPT	<b>0.035</b>	0	-	0
Pmed32	700	10	9297	7.3E+21	1.43	OPT	2	-	-	OPT	<b>0.224</b>	0	-	0
Pmed33	700	70	4700	3.37E+97	10.00	OPT	<b>6.8</b>	-	-	OPT	11.73	0	-	0
Pmed34	700	140	3013	5E+150	20.00	3015	9.8	-	-	3014	39.94	0.066	-	0.033
Pmed35	800	5	10400	2.7E+12	0.63	OPT	2.3	-	-	OPT	<b>0.048</b>	0	-	0
Pmed36	800	10	9934	2.8E+22	1.25	OPT	2.8	-	-	OPT	<b>0.232</b>	0	-	0
Pmed37	800	80	5057	4.1E+111	10.00	5058	11.4	-	-	5058	33.43	0.02	-	0.02
Pmed38	900	5	11060	4.87E+12	0.56	OPT	4.3	-	-	OPT	<b>0.104</b>	0	-	0
Pmed39	900	10	9423	9.14E+22	1.11	OPT	4	-	-	OPT	<b>0.256</b>	0	-	0
Pmed40	900	90	5128	5.1E+125	10.00	5133	19.9	-	-	5130	112.53	0.098	-	0.039
Average results Pmed1-15							0.2s		60.1s		1.35s	0.0154	0.007	0
Average results Pmed1-40							2.7s		-		17.7s	0.0360	-	0.002
Number of problems solved optimally							28		14		37			

#### IV. DISCUSSION

##### A. Proposed GA vs ADE GA [6]

The GA proposed in this study achieved OPT solutions in 37 of the 40 PMPs shown in Table IV. The non-OPT solutions present a mean deviation of 0.002% from OPT solution, corresponding to a mean time of 17.7s. ADE GA [6] presents OPT solutions in 28 of the 40 PMPs. The non-OPT solutions present a mean deviation of 0.036% from OPT solutions, corresponding to a mean time of 2.7s in a CPU Core I7 7500U at 1.8GHz.

Considering the 28 PMPs that both methods achieved OPT solutions, the proposed GA and the ADE GA [6] achieved best results in 21 and 7 of them, respectively. In the 7 PMPs that ADE GA [6] achieved best results, 6 of them occurred between Pmed1 and Pmed20. This range corresponds to less complex problems. To evaluate the performance difference between the two methods, we applied a Qui-Square test in the following 2x2 contingency table: [21 7; 7 21], and found  $\chi^2 = 14$ . For 1 degree of freedom, and a significance level of 99%, the critical level is  $t_c = 6.63$ . As  $\chi^2 > t_c$ , the difference between the proposed GA algorithm and ADE GA [6] is statistically significant.

From Table IV we also observe that when the ratio  $p/N$  increases, ADE GA [6] presents results significantly lower than the results obtained in this study. In the range Pmed21 to Pmed40, ADE GA [6] achieved OPT solutions in 10 of the 20 PMPs, with mean deviation of 0.046% from the OPT solutions, while the GA proposed in this study achieved OPT solutions in 17 of the PMPs, with mean deviation of 0.004% from the OPT solutions. For the instances Pmed5, Pmed10, Pmed15, Pmed20, Pmed25 and Pmed30, in which the ratio  $p/N$  is around 33%, the GA proposed in this study found all the OPT solutions, while ADE GA [6] found solutions with mean deviation of 0.18% from the OPT solutions. We believe that, for more complex PMP problems ( $N > 900$ ), the GA algorithm proposed in this study would obtain better values than ADE GA [6].

##### B. Proposed GA vs GA proposed in [7]

The GA proposed in [7] obtained solutions only for problems in the range Pmed1 to Pmed15. In this range, it obtained OPT solution in 14 PMPs, with a deviation of 0.07% from the OPT solution. The GA proposed in this study obtained OPT solutions in all this range.

Table IV shows that the GA proposed in this study converged in a shorter time than GA proposed in [7]. The last one is 44 times slower. This result suggests that the centroid technique used for population initialization in [7] as well as the continuous population updating have a negative impact in convergence time of the GA algorithm, making it slower.

#### V. CONCLUSION

A customization of GA operators for solving the  $p$ -median problem is proposed in this study. When applied to solve the

PMPs of OR-library, the proposed algorithm found OPT solutions in 37 of 40 PMPs, with a mean deviation of 0.002% and with a mean time of 17.7s.

#### ACKNOWLEDGMENTS

This research was financially supported by Samsung *Electronica da Amazonia Ltda*, under the terms of the Brazilian Federal Law number 8.387/91, through an agreement signed with Center for R&D in Electronic and Information Technology- CETELI/UFAM.

#### REFERENCES

- [1] O. Kariv; S.L. Hakimi. The  $p$ -median problems. In: An Algorithmic Approach to Network Location Problems. SIAM Journal on Applied Mathematics, 1274, Real World Applications. Philadelphia, 37, 539-560, 1979
- [2] R. Whitaker. A fast algorithm for the greedy interchange for large-scale clustering and median location problems. INFOR 21, 95-108, 1983
- [3] C. Beltran, C. Tadonki, J. Vial. Solving the  $p$ -median problem with a semi-lagrangian relaxation, Logilab Report, HEC, University of Geneva, Switzerland, 2004
- [4] F. Chiyoshi, R.D. Galvão. A statistical analysis of simulated annealing applied to the  $p$ -median problem. Annals of Operational Research 96:61-74, 2000. doi: 10.1023/A:1018982914742
- [5] S. Salhi. Defining tabu list size and aspiration criterion within tabu search methods. Computers and Operations Research 29, 67-86, 2002. doi: 10.1016/S0305-0548(00)00062-9
- [6] O. Alp, E. Erkut, Z. Drezner. An efficient genetic algorithm for the  $p$ -median problem. Annals Operational Research 122:21-42, 2003. doi: 10.1023/A:1026130003508
- [7] S. Satoglu, M. Oksuz, G. Kayakutlu, K. Buyukozkan. A genetic algorithm for the  $p$ -Median facility location problem. GJCI2016 – Global Joint Conference on industrial engineering, Istanbul, 2016.
- [8] J. E. Beasley. OR-library: distributing test problems by electronic mail. Journal of Operations Research Society 41:1069-1072, 1990. doi: 10.2307/2582903
- [9] G. Reinelt. TSLIB – a traveling salesman library. ORSA Journal of Computing, 3, pp. 376-384, 1991. doi: 10.1287/ijoc.3.4.376
- [10] H. Chen, N.S. Flann, D.W. Watson. Parallel genetic simulated annealing: A massively parallel SIMD approach. IEEE Transactions of Parallel Distributed Computation, 9 (Feb. 1998), pp. 126-136, 1998. doi: 10.1109/71.663870
- [11] Z. Drezner, J. Brinberg, N. Mladenovic, S. Salhi. New heuristic algorithms for solving the planar  $p$ -median problem. Comp. Operations Research, 62, pp. 296-304, 2015. doi: 10.1016/j.cor.2014.05.010
- [12] D. F. Albdaiwi, H.h. AboelFotoh. A GPU-based genetic algorithm for the  $p$ -median problem, Journal of Supercomputing, 73, pp 4221-4244, 2010. doi: 10.1007/s11227-017-2006-x
- [13] J. A. Moreno-Perez, J. M. Moreno-Vega, N. Mladenovic, Tabu Search and Simulated Annealing in  $p$ -median Problems. Talk at the Canadian Operational Research Society Conference, Montreal, 1994.
- [14] J. E. Beasley, 'OR-library', 1985. [Online]. Available: <http://people.brunel.ac.uk/~mastjbjeb/orlib/pmedinfo.html>. [Accessed: 04-Jul-2019]
- [15] D. Corus and P. S. Oliveto. Standard steady state genetic algorithms can hillclimb faster than mutation-only evolutionary algorithms. IEEE Tran. on Evolut. Comp., 2017. doi: 10.1109/TEVC.2017.2745715
- [16] J. Holland. Adaption in natural and artificial systems. The University of Michigan Press, Ann Arbor, 1975.
- [17] M. Vavouras, K. Papadimitriou, I. Papaefstathiou., High-speed FPGA-based implementations of a genetic algorithm, in: International Symposium on Systems, Architectures, Modeling, and Simulation, (IEEE2009), pp. 9-16, 2009.
- [18] K. Deliparaschos.; G. Doyamis, S. Tzafestas. A parameterised genetic algorithm IP core: FPGA design, implementation and performance evaluation Int. Journal of Electronics, 95, pp. 1149-1166, 2008.
- [19] R. P. Weicker, "Dhrystone: a synthetic systems programming benchmark," Communications of the ACM, vol. 27, no. 10, pp. 1013-1030, Oct 1984. 41.



# An algorithm for 1-space bounded cube packing

Łukasz Zielonka

Institute of Mathematics and Physics  
UTP University of Science and Technology  
Al. Prof. S. Kaliskiego 7, 85-789 Bydgoszcz, Poland  
Email: Lukasz.Zielonka@utp.edu.pl

**Abstract**—In this paper, we present a 1-space bounded cube packing algorithm with asymptotic competitive ratio 10.872.

**Index Terms**—Online algorithms, bin packing, cube, one-space bounded

## I. INTRODUCTION

IN THE bin packing problem, we receive a sequence of items of different sizes that must be packed into a finite number of bins in a way that minimizes the number of bins used. When all the items are accessible, the packing method is called *offline*. The packing method is called *online*, when items arrive one by one and each item has to be packed irrevocably into a bin before the next item is presented.

In the online version of packing a crucial parameter is the number of bins available for packing, i.e., *active bins*. Each incoming item is packed into one of the active bins; the remaining bins are not available at this moment. If we close one of the current active bins, we open a new active bin. Once an active bin has been closed, it can never become active again. When the method allows at most  $t$  active bins at the same time, it is called *t-space bounded*. Unbounded space model does not impose any limits on the number of active bins. It is natural to expect a packing method to be less efficient with fewer number of active bins. In this paper, we study 1-space bounded 3-dimensional cube packing.

Let  $S$  be a sequence of cubes. Denote by  $A(S)$  the number of bins used by the algorithm  $A$  to pack items from  $S$ . Furthermore, denote by  $OPT(S)$  the minimum possible number of bins used to pack items from  $S$  by the optimal offline algorithm. By the asymptotic competitive ratio for the algorithm  $A$  we mean:

$$R_A^\infty = \limsup_{n \rightarrow \infty} \sup_S \left\{ \frac{A(S)}{OPT(S)} \mid OPT(S) = n \right\}.$$

### A. Related work

The one-dimensional case of the space bounded bin packing problem has been extensively studied and the best possible algorithms are known: the Next-Fit algorithm [5] for the one-space bounded model and the Harmonic algorithm [6] when the number of active bins goes to infinity. The questions concerning  $t$ -space bounded  $d$ -dimensional packing ( $d \geq 2$ ) have been studied in a number of papers. For large number of active bins, Epstein and van Stee [1] presented a  $(\Pi_\infty)^d$ -competitive space bounded algorithm, where  $\Pi_\infty \approx 1.69103$  is the competitive ratio of the one-dimensional algorithm

Harmonic. Algorithms for 2-dimensional bin packing with only one active bin were explored for the first time in [8], where the authors give 8.84-competitive algorithm for 2-dimensional bin packing. An improved result of that case can be found in the paper [7], where a 5.155-competitive method is presented. The last article also contains an algorithm for packing squares with competitive ratio at most 4.5. In [4], a 4.84-competitive 1-space bounded 2-dimensional bin packing algorithm was presented. Grzegorek and Januszewski [3] presented a  $3.5^d$ -competitive as well as a  $12 \cdot 3^d$ -competitive online  $d$ -dimensional hyperbox packing algorithm with one active bin. The  $d$ -dimensional case of 1-space bounded hypercube packing was discussed in [9], where a  $2^{d+1}$ -competitive algorithm was described. The aim of this paper is to improve the upper bound ( $2^{3+1}$ ) in the 3-dimensional case. We present 10.872-competitive 1-space bounded cube packing algorithm.

### B. Our results

The algorithm presented in this article considers packing items (cubes of edges not greater than 1) into one active cube of edge 1. The main packing method is a bit like the classic computer game Tetris. The packing method which we describe is similar to the method presented by Grzegorek and Januszewski in [2]. The algorithm distinguishes types of items what determines a method for packing a specific item in a bin. Items that are considered big enough are packed from top to bottom. Different types of small items are packed from bottom upwards. The algorithm handles small items in a Tetris manner: to determine a place to pack an item a part of a bin is temporarily divided into congruent cuboids of appropriate size. Then an item is packed as low as possible inside a carefully chosen cuboid.

In Section II we give a 1-space bounded cube packing algorithm with the ratio 10.872.

## II. THE one-space-ALGORITHM

Let  $S$  be a sequence of cubes  $Q_1, Q_2, \dots$ . Denote by  $a_i$  the edge length of  $Q_i$ .

- an item  $Q_i$  is *huge*, provided  $a_i > 1/2$ ;
- an item  $Q_i$  is *big*, provided  $1/4 < a_i \leq 1/2$ ;
- an item  $Q_i$  is *small*, provided  $a_i \leq 1/4$ ; a small item  $Q_i$  is of *type k* provided  $2^{-k-1} < a_i \leq 2^{-k}$ .

Let  $\mathcal{B}$  be the active bin. To shorten the notation, a cuboid whose edges have lengths  $a \times a \times b$  will be called an  $(a, b)$ -cuboid.

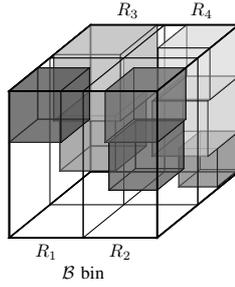


Fig. 1. Big items – the darker an item's colour, the later it arrived

### A. Description of the one-space-algorithm

- In packing items we distinguish coloured and white (not coloured) space. Items are placed only in the white space. Each newly opened bin is white.
- We divide each freshly opened bin into  $(1/2, 1)$ -cuboids. These cuboids are named  $R_1, R_2, R_3, R_4$  in an arbitrary order.
- Huge items (edge  $> 1/2$ ) are packed alone into a bin, i.e., if  $Q_i$  is huge, then we close the active bin and open a new bin to pack this item. After packing  $Q_i$  we close the bin and open a new active bin.
- If  $Q_i$  is big ( $1/4 < \text{edge} \leq 1/2$ ) we find the highest indexed  $R_j$  such that  $Q_i$  can be packed into it. We pack  $Q_i$  into  $R_j$  along the edge of  $\mathcal{B}$  as high as it is possible (see Figs. 1 and 3). If such a packing is not possible, we close the active bin, open a new active bin and pack  $Q_i$  into it.
- If  $Q_i$  is a small item of type  $k$  ( $2^{-k-1} < \text{edge} \leq 2^{-k}$ ) (see Figs. 2 and 3) we find the lowest indexed  $R_j$  such that  $Q_i$  can be packed into it. Since  $j$  is fixed now, we will write  $R$  instead of  $R_j$ .

When a big item is packed, it colours the space covered by itself.

We temporarily divide  $R$  into  $(2^{-k}, 1)$ -cuboids called  $R(1), \dots, R(4^{k-1})$ . Denote by  $t(n)$  the distance between the top of  $R(n)$  and the top of the topmost item packed in  $R(n)$  for  $n = 1, \dots, 4^{k-1}$  (see Fig. 5, right) and let  $\eta$  be an integer such that  $t(\eta) = \max\{t(1), \dots, t(4^{k-1})\}$ . We pack  $Q_i$  into  $R(\eta)$  as low as possible. The result of packing  $Q_i$  is the colouring of the  $(2^{-k}, 1 - t(\eta) + a_i)$ -cuboid contained in the bottom of  $R(\eta)$  (see Fig. 5, right, where  $\eta = 2$  before  $Q_{14}$  was packing).

If such a packing is not possible, then we close the active bin and open a new active bin to pack  $Q_i$ .

### B. Competitive ratio

Let  $P_j$  for  $j = 1, \dots, 16$  be  $(1/4, 1)$ -cuboids with pairwise disjoint interiors. Each cuboid  $R_i$  for  $i \in \{1, 2, 3, 4\}$  is divided into four cuboids  $P_{4i-3}, \dots, P_{4i}$  (see Fig. 4).

**Lemma 1.** Assume that only small items were packed into  $\mathcal{B}$ . Assume that  $j \in \{1, 2, \dots, 16\}$ . Denote by  $n$  the number of items packed into  $P_j$  and by  $t_n$  the distance between the

bottom of  $\mathcal{B}$  and the top of the topmost item packed into  $P_j$ . The total volume  $v_n$  of small items packed into  $P_j$  is greater than

$$f(t_n) = \frac{19}{2048} \cdot t_n - \frac{13}{16384}.$$

Moreover, if the topmost packed item is of type 2, then

$$v_n > f_+(t_n) = \frac{19}{2048} \cdot t_n.$$

*Proof.* Without loss of generality we can assume that  $P_j = [0, 1/4] \times [0, 1/4] \times [0, 1]$ . We will prove the result using induction over the number  $n$  of packed items.

First assume that only one item  $Q_b$  was packed into  $P_j$ . Obviously,  $t_1 = a_b$ . Let

$$\varphi(a) = a^3 - \frac{19}{2048}a.$$

The function  $\varphi(a)$  for  $a > 0$  has a minimum at

$$a_0 = \sqrt{\frac{19}{6144}}.$$

A computation shows that

$$\varphi(a_0) > -\frac{1}{2} \cdot \frac{13}{16384} \quad (1)$$

(this lower bound will be useful in the last part of the proof). We get

$$v_1 = a_b^3 > \frac{19}{2048} \cdot t_1 - \frac{1}{2} \cdot \frac{13}{8192} = f(t_1).$$

Moreover, if  $1/8 < a_b \leq 1/4$ , then  $v_1 = a_b^3 > \frac{19}{2048}a_b = f_+(t_1)$ .

Now assume that the statement holds for at most  $n$  items packed into  $P_j$  (this is our inductive assumption). Let  $Q_u$  be the  $(n+1)$ st item packed into  $P_j$  and let  $t_{n+1}$  be the distance between the bottom of  $P_j$  and the top of the topmost item (from among  $n+1$  items  $Q_b, \dots, Q_u$ ) packed into  $P_j$ .

If  $a_u > 1/8$ , then  $t_{n+1} = t_n + a_u$ . Using the inductive assumption,

$$v_{n+1} = v_n + a_u^3 > f(t_n) + a_u^3 = \frac{19}{2048} \cdot t_n - \frac{13}{16384} + a_u^3.$$

Since

$$\varphi'(a) = 3a^2 - \frac{19}{2048} > 3 \cdot \frac{1}{64} - \frac{19}{2048} > 0$$

for  $a > 1/8$ , we get

$$\varphi(a) > \varphi\left(\frac{1}{8}\right) = \frac{13}{16384}$$

for  $a > 1/8$ . Consequently,

$$\begin{aligned} v_{n+1} &> f(t_n) + a_u^3 = \frac{19}{2048}(t_n + a_u) + \varphi(a_u) - \frac{13}{16384} \\ &\geq \frac{19}{2048}(t_n + a_u) = f_+(t_n + a_u) = f_+(t_{n+1}). \end{aligned}$$

Finally, consider the case when  $a_u \leq 1/8$ . First, we choose the topmost packed item  $Q^1$  with edge greater than  $1/8$  and denote by  $\tau$  the distance between the bottom of  $P_j$  and the top of  $Q^1$  (see Fig. 6, left). If there is no such item, then we

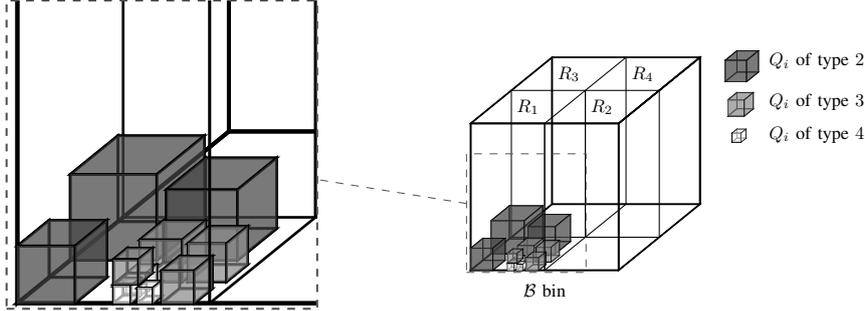


Fig. 2. Small items

take  $\tau = 0$ . The total volume of items packed up to  $\tau$ , by the inductive assumption, is not smaller than  $f_+(\tau)$ . Above  $Q^1$  we divide  $P_j$  into four  $(1/8, 1 - \tau)$ -cuboids  $P_j^1, P_j^2, P_j^3, P_j^4$ . Denote by  $Q_1^l, \dots, Q_{v_l}^l$  the items from among  $Q_b, \dots, Q_{u-1}$  packed into  $P_j^l$  above  $Q^1$  (if any) for each  $l = 1, 2, 3, 4$ . Moreover, denote by  $t_n^l$  the distance between the bottom of  $P_j$  and the top of the topmost item from among  $Q_b, \dots, Q_{u-1}$  packed into  $P_j^l$  and let  $t_n^* = \min(t_n^1, t_n^2, t_n^3, t_n^4)$  (see Fig. 6, right). Clearly,  $t_n^* \geq \tau$  and  $t_n^* \leq t_n$ .

If  $t_n^* + a_u \leq t_n$ , then  $t_{n+1} = t_n$ . Consequently,

$$v_{n+1} \geq v_n + a_u^3 = f(t_n) + a_u^3 = f(t_{n+1}) + a_u^3 > f(t_{n+1}).$$

If  $t_n^* + a_u > t_n$ , then  $t_{n+1} = t_n^* + a_u$ . Items  $Q_1^l, \dots, Q_{v_l}^l$  were packed into  $(1/8, t_n^l - \tau)$ -cuboid  $P_j^l$ . Let  $h(P_j^l) = [0, 1/4] \times [0, 1/4] \times [0, 2t_n^l - 2\tau]$  be the image of  $P_j^l$  in a homothety  $h$  of ratio 2. By the inductive assumption, the total volume of cubes  $h(Q_1^l), \dots, h(Q_{v_l}^l)$  is not smaller than  $\frac{19}{2048}(2t_n^l - 2\tau) - \frac{13}{16384} = f(2t_n^l - 2\tau)$ . Since the volume of each  $h(Q_i^l)$  is 8 times greater than the volume of  $Q_i^l$ , it follows that the total volume of cubes  $Q_1^l, \dots, Q_{v_l}^l$  is not smaller than  $\frac{1}{8}f(2t_n^l - 2\tau)$ .

Consequently,

$$\begin{aligned} v_{n+1} &\geq f_+(\tau) + 4 \cdot \frac{1}{8} f(2t_n^* - 2\tau) + a_u^3 \\ &= a_u^3 + \frac{19}{2048} t_n^* - \frac{1}{2} \cdot \frac{13}{16384}. \end{aligned}$$

By (1) we know that

$$\varphi(a_0) > -\frac{1}{2} \cdot \frac{13}{16384}.$$

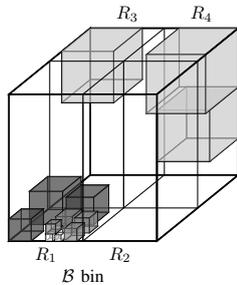


Fig. 3. one-space-algorithm

Consequently,

$$\begin{aligned} v_{n+1} &\geq \varphi(a_u) + \frac{19}{2048}(t_n^* + a_u) - \frac{1}{2} \cdot \frac{13}{16384} \\ &> \frac{19}{2048}(t_n^* + a_u) - \frac{13}{16384} = f(t_{n+1}). \end{aligned}$$

□

**Lemma 2.** Define  $V_3 = 101/1024$ . Let  $S$  be a finite sequence of cubes and let  $\nu$  be the number of bins used to pack items from  $S$  by the one-space-algorithm. Moreover, let  $m$  be the number of huge items in  $S$ . The total volume of items in  $S$  is greater than  $2^{-3} \cdot m + V_3(\nu - 2m - 1)$ .

*Proof.* Among  $\nu$  bins used to pack items from  $S$  by the one-space-algorithm the first  $\nu - 1$  bins will be called full. Let  $Q_z$  be the first item from  $S$  which cannot be packed into a full bin  $\mathcal{B}$  by the one-space-algorithm. Clearly,  $Q_z$  is the first item packed into the next bin.

Denote by  $v_{\mathcal{B}}$  the sum of volumes of items packed into  $\mathcal{B}$ .

If the incoming item  $Q_z$  is huge, then the average occupation ratio in both bins  $\mathcal{B}_j$  and the next bin  $\mathcal{B}_{j+1}$  into which  $Q_z$  was packed is greater than  $1/2^4$ . Obviously, there are  $2m$  such bins.

It is possible that the last bin is almost empty.

To prove Lemma 2 it suffices to show that if  $Q_z$  is not huge and if no huge item was packed into  $\mathcal{B}$ , then  $v_{\mathcal{B}} > V_3$  (the number of such bins equals  $\nu - 2m - 1$ ).

*Case 1:  $Q_z$  is small and all items packed into  $\mathcal{B}$  are small.*

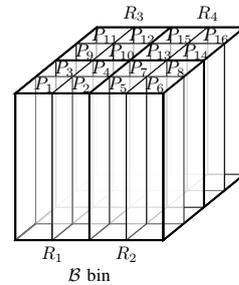


Fig. 4.  $(1/4, 1)$ -cuboids  $P_j$

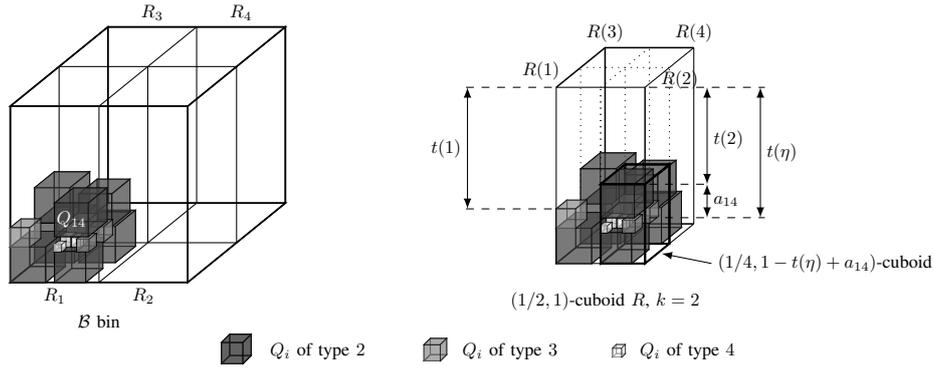


Fig. 5. Packing small items into  $(2^{-k}, 1)$ -cuboids

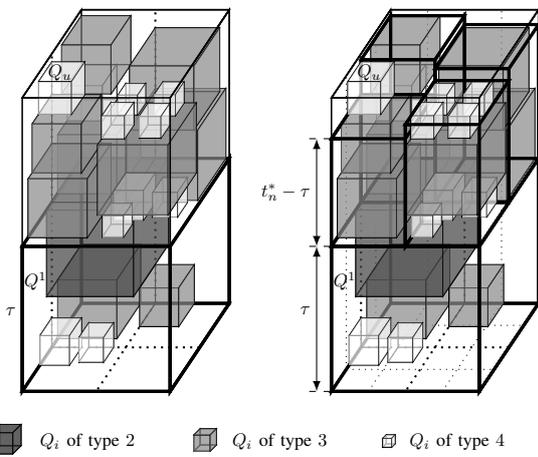


Fig. 6. The division

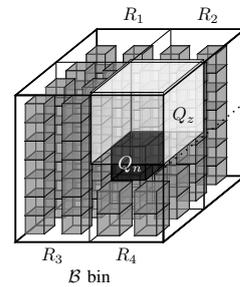


Fig. 7. Case 3

Since  $a_z \leq 1/4$ , it follows that each  $P_i$  is packed up to height at least  $3/4$ . By Lemma 1 we deduce that

$$v_B > 4^2 f\left(\frac{3}{4}\right) = 16 \cdot \left(\frac{19}{2048} \cdot \frac{3}{4} - \frac{13}{4 \cdot 16384}\right) = V_3.$$

Case 2:  $Q_z$  is small and a big item was packed into  $\mathcal{B}$ .

The volume of a big item  $Q_b$  with edge  $t$  is equal to  $t^3 > t \cdot \left(\frac{1}{4}\right)^2$ . In considerations presented in Case 1 we accept that the total volume of small items packed into  $R_j$  up to height  $t$  equals  $4f(t)$ . It is easy to see that

$$4f(t) < \frac{1}{16} \cdot t.$$

As a consequence,  $v_B > V_3$ .

Case 3:  $Q_z$  is a big item and all items packed into  $\mathcal{B}$  are small

Assume that there is  $(2^{-2}, 1)$ -cuboid  $R_j(n)$  ( $j \in \{1, 2, 3\}$ ,  $n \in \{1, 2, 3, 4\}$ ) such that the distance between its top and the top of the topmost item packed into it is greater than  $1/8$  and denote by  $R_+$  first such cuboid. The total volume of items packed into  $R_+$  is greater than  $f(3/4)$ . The total volume of items packed into each cuboid preceding  $R_+$  is greater than  $f(7/8)$ . The total volume of items packed into

each of remaining cuboids is greater than  $\frac{3}{4} \cdot \frac{1}{8^2} > f(7/8)$  (in such a cuboid only items greater than  $1/8$  were packed).

Denote by  $Q_n$  the topmost small item packed in  $R_4$  (as in Fig. 7). Since  $a_z \leq 1/2$  and  $Q_z$  cannot be packed in  $R_4$ , it follows that

$$v_B > (16 - 5)f\left(\frac{7}{8}\right) + f\left(\frac{3}{4}\right) + 4f\left(\frac{1}{2} - a_n\right) + a_n^3.$$

Denote by  $\gamma(a_n)$  the function on the right-hand side of this formula. This function for positive  $a$  has a minimum at  $a_0 = \sqrt{\frac{19}{1536}}$ .

A computation shows that  $\gamma(a_0) > V_3$ . Consequently,  $v_B > V_3$ .

If there is no  $(2^{-2}, 1)$ -cuboid  $R_j(n)$  ( $j \in \{1, 2, 3\}$ ,  $n \in \{1, 2, 3, 4\}$ ) such that the distance between its top and the top of the topmost item packed into it is greater than  $1/8$ , then

$$v_B > (16 - 4)f\left(\frac{7}{8}\right) + 4f\left(\frac{1}{2} - a_n\right) + a_n^3.$$

Since  $f(7/8) > f(3/4)$ , we get  $v_B > \gamma(a_0) > V_3$ .

Case 4:  $Q_z$  is big and a big item was packed into  $\mathcal{B}$

Similarly as in Case 2 we get

$$4f(t) < t^3.$$

We deduce by Case 3 that  $v_B > V_3$ . □

**Theorem 1.** *The asymptotic competitive ratio for the one-space-algorithm is not greater than  $1098/101 \approx 10.8713$ .*

*Proof.* Let  $S$  be a sequence of items of total volume  $v$ , let  $m$  denote the number of huge items in  $S$  and let  $\mu$  be the number of bins used to pack items from  $S$  using the one-space-algorithm. Obviously,  $OPT(S) \geq v$  as well as  $OPT(S) \geq m$ .

By Lemma 2 we get  $v > \frac{1}{2^3} \cdot m + V_3 \cdot (\mu - 2m - 1)$ , i.e.,

$$\mu < \frac{v}{V_3} + m \left( 2 - \frac{1}{2^3 V_3} \right) + 1.$$

It is easy to check that  $2 - \frac{1}{8V_3} > 0$ .

If  $m < v$ , then

$$\frac{\mu}{OPT(S)} \leq \frac{\mu}{v} < \frac{\frac{v}{V_3} + v \left( 2 - \frac{1}{2^3 V_3} \right) + 1}{v} = \frac{2^3 - 1}{2^3 V_3} + 2 + \frac{1}{v}.$$

If  $v \leq m$ , then

$$\frac{\mu}{OPT(S)} \leq \frac{\mu}{m} \leq \frac{\frac{m}{V_3} + m \left( 2 - \frac{1}{2^3 V_3} \right) + 1}{m} = \frac{2^3 - 1}{2^3 V_3} + 2 + \frac{1}{m}.$$

Consequently, the asymptotic competitive ratio for the one-space-algorithm is not greater than

$$\frac{7}{8} \cdot \frac{1024}{101} + 2 = \frac{1098}{101} < 10.872.$$

□

## REFERENCES

- [1] L. Epstein and R. van Stee. Optimal online algorithms for multidimensional packing problems. *SIAM Journal on Computing*, 35(2):431–448, 2005.
- [2] P. Grzegorek and J. Januszewski. A note on one-space bounded square packing. *Information Processing Letters*, 115(11):872–876, 2015.
- [3] P. Grzegorek, J. Januszewski. Drawer algorithms for 1-space bounded multidimensional hyperbox packing. *Journal of Combinatorial Optimization*, 37(3): 1011–1044, 2019.
- [4] J. Januszewski and Ł. Zielonka. Online packing of rectangular items into square bins. In R. Solis-Oba and R. Fleischer, editors, *Approximation and Online Algorithms. WAOA 2017*, volume 10787 of *Lecture Notes in Computer Science*, pages 147–163, Cham, 2018. Springer.
- [5] D. S. Johnson. Fast algorithms for bin packing. *Journal of Computer and System Sciences*, 8(3):272–314, 1974.
- [6] C.-C. Lee and D.-T. Lee. A simple on-line bin-packing algorithm. *J. ACM*, 32(3):562–572, July 1985.
- [7] Y. Zhang, J. Chen, F. Y. L. Chin, X. Han, H.-F. Ting, and Y. H. Tsin. Improved online algorithms for 1-space bounded 2-dimensional bin packing. In O. Cheong, K.-Y. Chwa, and K. Park, editors, *Algorithms and Computation*, pages 242–253, Berlin, Heidelberg, 2010. Springer.
- [8] Y. Zhang, F. Y. L. Chin, and H.-F. Ting. One-space bounded algorithms for two-dimensional bin packing. *International Journal of Foundations of Computer Science*, 21(06):875–891, 2010.
- [9] Y. Zhang, F. Y. L. Chin, H.-F. Ting, and X. Han. Online algorithms for 1-space bounded multi dimensional bin packing and hypercube packing. *Journal of Combinatorial Optimization*, 26(2):223–236, 2013.



# Ant Colony Optimization Algorithm for Workforce Planning: Influence of the Evaporation Parameter

Stefka Fidanova  
IICT, BAS  
Sofia, Bulgaria  
E-mail: stefka@parallel.bas.bg

Gabriel Luque  
DLCS University of Málaga  
29071 Málaga, Spain  
E-mail: gabriel@lcc.uma.es

Olympia Roeva  
IBPhBME, BAS  
Sofia, Bulgaria  
E-mail: olympia@biomed.bas.bg

Maria Ganzha  
SRI, PAS  
Warsaw, Poland  
E-mail: maria.ganzha@ibspan.waw.pl

**Abstract**—Optimization of the production process is important for every factory or organization. The better organization can be done by optimization of the workforce planing. The main goal is decreasing the assignment cost of the workers with the help of which, the work will be done. The problem is NP-hard, therefore it can be solved with algorithms coming from artificial intelligence. The problem is to select employers and to assign them to the jobs to be performed. The constraints of this problem are very strong and for the algorithms is difficult to find feasible solutions. We apply Ant Colony Optimization Algorithm to solve the problem. We investigate the algorithm performance according evaporation parameter. The aim is to find the best parameter setting.

**Index Terms**—Workforce Planning, Ant Colony Optimization, Metaheuristics, Evaporation parameter

## I. INTRODUCTION

**T**HE WORKFORCE planning is a very important decision making problem for branches of the industry. It plays an important role in human resource management. It includes multiple level of complexity, therefore it is a hard optimization problem (NP-hard). The problem can be divided in to two parts: selection and assignment. The first part is selection of employers from the set of available workers. The second part is assignment of the selected workers to jobs, which the worker will perform. The goal is to carry out the work requirements minimizing assignment cost.

As we have noted above the problem is very hard optimization problem and is impossible to be solved with exact methods or traditional numerical methods for instances with realistic size. These types of methods can be applied only on simplified variants of the problem. A deterministic version of workforce planing problem is studied in [12], [18]. In [12] the workforce planning is reformulated as mixed integer programming. It is shown that the mixed integer program is much easier to solve the problem than the non-linear program. In [18] the model includes workers differences and the possibility of workers training and upgrading. A variant with random demands of the problem is considered in [4], [19]. Two stage program of scheduling and allocating with

random demands is proposed in [4]. Other variant of the problem is to include uncertainty [13], [15], [17], [24], [25]. A lot of authors skip some of the constraints to simplify the problem. Mixed linear programming is apply in [6] and in [19] is utilized decomposition method, but for the more complex non-linear workforce planning problems, the convex methods are not applicable.

Last decade, nature-inspired metaheuristic methods receive more and more attention, because they can find close to optimal solutions even for large-scale difficult problems [2], [16], [20], [22], [23]. In the literature can be found various metaheuristic algorithms solving workforce planning problem. They include genetic algorithm [1], [14], memetic algorithm [21], scatter search [1] etc.

Ant Colony Optimization (ACO) algorithm is proved to be very effective solving various complex optimization problems [7], [11]. In our previous work [8], [9] we propose ACO algorithm for workforce planning. We have considered the variant of the workforce planning problem proposed in [1]. Current paper is the continuation of [8] and further develops the ideas behind [8]. We investigate the influence of evaporation parameter on algorithm performance. The aim is to find the best parameter setting.

The rest of the paper is organized as follows. In Section 2 the mathematical description of the problem is presented. In Section 3 ACO algorithm for workforce planing problem is described. Section 4 shows computational results, comparisons and discussion. A conclusion and directions for future work are done in Section 5.

## II. WORKFORCE PLANNING PROBLEM

In this paper we consider the workforce planning problem proposed in [1] and [10]. The set of jobs  $J = \{1, \dots, m\}$  need to be completed during a fixed period of time. The job  $j$  requires  $d_j$  hours to be finished.  $I = \{1, \dots, n\}$  is the set of workers, candidates to be assigned. Every worker must perform every of assigned to him job minimum  $h_{min}$  hours can work in efficient way. The worker  $i$  is available  $s_i$  hours.

One worker can be assigned to maximum  $j_{max}$  jobs. The set  $A_i$  shows the jobs, that worker  $i$  is qualified. Maximum  $t$  workers can be assigned during the planed period, or at most  $t$  workers may be selected from the set  $I$  of workers. The selected workers need to be capable to complete all the jobs they are assigned. The goal is to find feasible solution, that optimizes the objective function.

The cost of assigning the worker  $i$  to the job  $j$  is  $c_{ij}$ . The mathematical model of the workforce planing problem is described as follows:

$$x_{ij} = \begin{cases} 1 & \text{if the worker } i \text{ is assigned to job } j \\ 0 & \text{otherwise} \end{cases}$$

$$y_i = \begin{cases} 1 & \text{if worker } i \text{ is selected} \\ 0 & \text{otherwise} \end{cases}$$

$$z_{ij} = \text{number of hours that worker } i \\ \text{is assigned to perform job } j$$

$$Q_j = \text{set of workers qualified to perform job } j$$

$$\text{Minimize } \sum_{i \in I} \sum_{j \in A_i} c_{ij} \cdot x_{ij} \quad (1)$$

Subject to

$$\sum_{j \in A_i} z_{ij} \leq s_i \cdot y_i \quad i \in I \quad (2)$$

$$\sum_{i \in Q_j} z_{ij} \geq d_j \quad j \in J \quad (3)$$

$$\sum_{j \in A_i} x_{ij} \leq j_{max} \cdot y_i \quad i \in I \quad (4)$$

$$h_{min} \cdot x_{ij} \leq z_{ij} \leq s_i \cdot x_{ij} \quad i \in I, j \in A_i \quad (5)$$

$$\sum_{i \in I} y_i \leq t \quad (6)$$

$$x_{ij} \in \{0, 1\} \quad i \in I, j \in A_i$$

$$y_i \in \{0, 1\} \quad i \in I$$

$$z_{ij} \geq 0 \quad i \in I, j \in A_i$$

The objective function is the minimization of the total assignment cost. The number of hours for each selected worker is limited (inequality 2). The work must be done in full (inequality 3). The number of the jobs, that every worker can perform is limited (inequality 4). There is minimal number of hours that every job must be performed by every assigned worker can work efficiently (inequality 5). The number of assigned workers is limited (inequality 6).

This mathematical model can be used with other objectives too. If  $\tilde{c}_{ij}$  is the cost the worker  $i$  to performs the job  $j$  for one hour, than the objective function can minimize the cost of the hall jobs to be finished.

$$f(x) = \text{Min} \sum_{i \in I} \sum_{j \in A_i} \tilde{c}_{ij} \cdot x_{ij} \quad (7)$$

The preferences of the workers to the jobs can be included. In this case one of the variants of the objective function will be to maximize the satisfaction of the workers preferences.

### III. ANT COLONY OPTIMIZATION ALGORITHM

The ACO is a nature inspired methodology. It is a meta-heuristics, following the real ants behavior when looking for a food. Real ants use chemical substance, called pheromone, to mark their path ant can return back. An ant moves in random way and when it detects a previously laid pheromone it decides whether to follow it and reinforce it with a new added pheromone. Thus the more ants follow a given trail, the more attractive that trail becomes. There is evaporation in a nature and the pheromone evaporates during the time. Thus the pheromone level of not used and less used paths decreases and they become less desirable. In this way the nature prevents the ants to follow some wrong and useless path. The ants can find a shorter path between the source of the food and the nest by their collective intelligence.

#### A. Main ACO Algorithm

It is not practical to solve HP-hard problems with exact methods or traditional numerical methods when the problem is large. An option is to be applied some metaheuristics. The goal is to find a good solution for a reasonable computational resources like time and memory [5].

For a first time, ant behavior is used for solving optimization problems by Marco Dorigo [3]. Later some modifications are proposed, mainly in pheromone updating rules [5]. The basic in ACO methodology is the simulation of ants behavior. The problem is represented by graph. The solutions are represented by paths in a graph and the aim is to find shorter path corresponding to given constraints. The requirements of ACO algorithm are as follows:

- Appropriate representation of the problem by a graph;
- Appropriate pheromone placement on the nodes or on the arcs of the graph;
- Suitable problem-dependent heuristic function, which manage the ants to improve solutions;
- Pheromone updating rules;
- Transition probability rule, which specifies how to include new nodes in the partial solution;
- Appropriate algorithm parameters.

The transition probability  $P_{i,j}$ , is a product of the heuristic information  $\eta_{i,j}$  and the pheromone trail level  $\tau_{i,j}$  related to the move from node  $i$  to the node  $j$ , where  $i, j = 1, \dots, n$ .

$$P_{i,j} = \frac{\tau_{i,j}^a \cdot \eta_{i,j}^b}{\sum_{k \in Unused} \tau_{i,k}^a \cdot \eta_{i,k}^b}, \quad (8)$$

where  $Unused$  is the set of unused nodes of the graph.

The initial pheromone level is the same for all elements of the graph and is set to a positive constant value  $\tau_0$ ,  $0 < \tau_0 < 1$ .

After that at the end of the current iteration the ants update the pheromone level [5]. A node become more desirable if it accumulates more pheromone.

The main update rule for the pheromone is:

$$\tau_{i,j} \leftarrow \rho \cdot \tau_{i,j} + \Delta\tau_{i,j}, \quad (9)$$

where  $\rho$  decreases the value of the pheromone, which mimics evaporation in a nature.  $\Delta\tau_{i,j}$  is a new added pheromone, which is proportional to the quality of the solution. For measurement of the quality of the solution is used the value of the objective function of the ants solution.

The first node of the solution is randomly chosen. With the random start the search process is diversifying and the number of ants may be small according the number of the nodes of the graph and according other population based metaheuristic methods. The heuristic information represents the prior knowledge of the problem, which is used to better manage the algorithm performance. The pheromone is a global history of the ants to find optimal solution. It is a tool for concentration of the search around best so far solutions.

### B. Workforce Planing ACO

An important role for the successes of the ACO algorithm is the representation of the problem by graph. The graph we propose is 3 dimensional and the node  $(i, j, z)$  corresponds to worker  $i$  to be assigned to the job  $j$  for time  $z$ . When an ant begins their tour we generate three random numbers: the first random number is from the interval  $[0, \dots, n]$  and corresponds to the worker we assign; the second random number is from the interval  $[0, \dots, m]$  and shows the job which this worker will perform. The third random number is from the interval  $[h_{min}, \min\{d_j, s_i\}]$  and shows number of hours worker  $i$  is assigned to performs job  $j$ . Next node is included in the solution, applying transition probability rule. We repeat this procedure till the solution is constructed.

The following heuristic information is applied:

$$\eta_{ijl} = \begin{cases} l/c_{ij} & l = z_{ij} \\ 0 & otherwise \end{cases} \quad (10)$$

By this heuristic information the most cheapest unassigned worker, is assigned as longer as possible. The node with highest probability from all possible nodes is chosen to be included in the partial solution. When there are more than one possibilities with the same probability, the next node is chosen in a random way between them.

When a new node is included we take in to account all constraints: how many workers are assigned till now; how many time slots every worker is assigned till now; how many time slots are assigned per job till now. If a move do not meets all constraints, the probability of this move is set to 0. The solution is constructed if there are not more possibilities for including new nodes (the transition probability is 0 for all possible moves). If the constructed solution is feasible the value of the objective function is the sum of the assignment cost of the assigned workers. When the constructed solution

TABLE I: Test instances characteristics

Parameters	Value
$n$	20
$m$	20
$t$	10
$s_i$	[50,70]
$j_{max}$	[3,5]
$h_{min}$	[10,15]

TABLE II: ACO parameter settings

Parameters	Value
Number of iterations	100
$\rho$	{0.1, 0.3, 0.5, 0.7, 0.9}
$\tau_0$	0.5
Number of ants	20
$a$	1
$b$	1

is not feasible, the value of the objective function is set to be equal to  $-1$ .

New pheromone is deposited only on the elements of feasible solutions. The deposited pheromone is proportional to the reciprocal value of the objective function.

$$\Delta\tau_{i,j} = \frac{\rho - 1}{f(x)} \quad (11)$$

Thus the nodes belonging to better solutions accumulate more pheromone than others and will be more attractive in the next iteration. The iteration best solution is compared with the global best solution and if on the current iteration the some of the ants achieves better solution it becomes the new global best. As end condition we use the number of iterations.

In this research we are concentrated on influence of the evaporation parameter on algorithm performance. We tested several values for this parameter and compare the number of needed iterations to find the best solution.

## IV. COMPUTATIONAL RESULTS AND DISCUSSION

In this section we tested our algorithm and evaporation parameter influence on 10 structured problems. The software, which realizes the algorithm is written in C and is run on Pentium desktop computer at 2.8 GHz with 4 GB of memory.

An artificially generated problem instances considered in [1] is used for the tests. The test instances characteristics are shown in Table I.

The parameter settings of our ACO algorithm is shown in Table II and are fixed experimentally after several runs of the algorithm.

In our previous work [8] we show that our ACO algorithm outperforms the genetic and scatter search algorithms proposed in [1]. We perform 30 independent runs with every one of the five values of the evaporation parameter, because the algorithm is stochastic and to guarantee the robustness of the

TABLE III: Evaporation parameter ranking

	$\rho = 0.1$	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$
first place	3 times	4 times	1 times	2 times	0 times
second place	2 times	3 times	3 times	2 times	1 times
third place	2 times	1 times	3 times	2 times	1 times
forth place	2 times	1 times	2 times	3 times	2 times
fifth place	1 times	1 times	1 times	1 times	6 times
ranking	26	22	29	31	43

average results. We apply ANOVA test for statistical analysis to guarantee the significance of the difference between the average results. We compare the average number of iterations needed to find the best result for every test problem. The needed number of iterations for every test problem can be very different, because the specificity of the tests. Therefore for comparison we use ranking as more representative. The algorithm with some fixed value for evaporation is on the first place, if it achieves the best solution with less average number of iterations over 30 runs, according other values and we assign to it 1, we assign 2 to the value on the second place, 3 to the value on the third place, 4 to the value of the forth place and 5 to the value with most number of iterations. On some cases can be assigned same numbers if the number of iterations to find the best solution is the same. We sum the ranking of the cases over all 10 test problems to find final ranking of the different values of the evaporation parameter.

Table III shows the ranking of the evaporation parameter. The less number of iterations is needed when the evaporation parameter is equal to 0.3. In this case the algorithm is on the first place four times, on the second place is 3 times and on the third, fourth and fifth is respectively one time. The worst results are achieved when the evaporation parameter is equal to 0.9. The results achieved when the evaporation parameter is 0.1 are a little bit worse than when the evaporation parameter is equal to 0.3. When the value of the evaporation parameter increase the achieved results are getting worse.

## V. CONCLUSION

In this paper we apply ACO algorithm to solve workforce planning problem. We are concentrated on the influence of the evaporation parameter on the algorithm performance, how many iterations are needed to find the best solution. We test the algorithm on 10 structured benchmark problems. The achieved results show that when the value of the evaporation parameter is small, the algorithm needs less number of iterations compared with high value of the evaporation parameter.

## ACKNOWLEDGMENT

Work presented here is partially supported by the National Scientific Fund of Bulgaria under grant DFNI DN12/5 “Efficient Stochastic Methods and Algorithms for Large-Scale Problems”, Grant No BG05M2OP001-1.001-0003, financed by the Science and Education for Smart Growth Operational Program and co-financed by the European Union through the

European structural and Investment funds, and by the Polish-Bulgarian collaborative grant “Practical aspects for scientific computing”.

## REFERENCES

- [1] Alba E., Luque G., Luna F., *Parallel Metaheuristics for Workforce Planning*, J. Mathematical Modelling and Algorithms, Vol. 6(3), Springer, 2007, 509-528. <https://doi.org/10.1007/s10852-007-9058-5>
- [2] Albayrak G., Özdemir İ., *A state of art review on metaheuristic methods in time-cost trade-off problems*, International Journal of Structural and Civil Engineering Research, Vol. 6(1), 2017, 30-34. <https://doi.org/10.18178/ijscer.6.1.30-34>
- [3] Bonabeau E., Dorigo M. and Theraulaz G., *Swarm Intelligence: From Natural to Artificial Systems*, New York, Oxford University Press, 1999.
- [4] Campbell G., *A two-stage stochastic program for scheduling and allocating cross-trained workers*, J. Operational Research Society 62(6), 2011, 1038-1047. <https://doi.org/10.1057/jors.2010.16>
- [5] Dorigo M., Stutzle T., *Ant Colony Optimization*, MIT Press, 2004.
- [6] Easton F., *Service completion estimates for cross-trained workforce schedules under uncertain attendance and demand*, Production and Operational Management 23(4), 2014, 660-675. <https://doi.org/10.1111/poms.12174>
- [7] Fidanova S., Roeva O., Paprzycki M., Gepner P., *InterCriteria Analysis of ACO Start Strategies*, Proceedings of the 2016 Federated Conference on Computer Science and Information Systems, 2016, 547-550. [https://doi.org/10.1007/978-3-319-99648-6\\_4](https://doi.org/10.1007/978-3-319-99648-6_4)
- [8] Fidanova S., Luquq G., Roeva O., Paprzycki M., Gepner P., *Ant Colony Optimization Algorithm for Workforce Planning*, FedCSIS'2017, IEEE Xplorer, IEEE catalog number CFP1585N-ART, 2017, 415-419. <https://doi.org/10.15439/2017F63>
- [9] Roeva O., Fidanova S., Luque G., Paprzycki M., Gepner P., *Hybrid Ant Colony Optimization Algorithm for Workforce Planning*, FedCSIS'2018, IEEE Xplorer, 2018, 233-236. <https://doi.org/10.15439/2018F47>
- [10] Glover F., Kochenberger G., Laguna M., Wubben, T. *Selection and assignment of a skilled workforce to meet job requirements in a fixed planning period*. In: MAEB'04, 2004, 636-641.
- [11] Grzybowska K., Kovács, G., *Sustainable Supply Chain - Supporting Tools*, Proceedings of the 2014 Federated Conference on Computer Science and Information Systems, Vol. 2, 2014, 1321-1329. <https://doi.org/10.15439/2014F75>
- [12] Hewitt M., Chacosky A., Grasman S., Thomas B., *Integer programming techniques for solving non-linear workforce planning models with learning*, European J of Operational Research 242(3), 2015, 942-950. <https://doi.org/10.1016/j.ejor.2014.10.060>
- [13] Hu K., Zhang X., Gen M., Jo J., *A new model for single machine scheduling with uncertain processing time*, J Intelligent Manufacturing, Vol 28(3), Springer, 2015, 717-725. <https://doi.org/10.1007/s10845-015-1033-9>
- [14] Li G., Jiang H., He T., *A genetic algorithm-based decomposition approach to solve an integrated equipment-workforce-service planning problem*, Omega, Vol. 50, Elsevier, 2015, 1-17. <https://doi.org/10.1016/j.omega.2014.07.003>
- [15] Li R., Liu G., *An uncertain goal programming model for machine scheduling problem*. J. Intelligent Manufacturing, Vol. 28(3), Springer, 2014, 689-694. <https://doi.org/10.1007/s10845-014-0982-8>
- [16] Mucherino A., Fidanova S., Ganzha M., *Introducing the environment in ant colony optimization*, Recent Advances in Computational Optimization, Studies in Computational Intelligence, Vol. 655, 2016, 147-158. [https://doi.org/10.1007/978-3-319-40132-4\\_9](https://doi.org/10.1007/978-3-319-40132-4_9)
- [17] Ning Y., Liu J., Yan L., *Uncertain aggregate production planning*, Soft Computing, Vol. 17(4), Springer, 2013, 617-624. <https://doi.org/10.1007/s00500-012-0931-4>
- [18] Othman M., Bhuiyan N., Gouw G., *Integrating workers' differences into workforce planning*, Computers and Industrial Engineering, Vol. 63(4), 2012, 1096-1106. <https://doi.org/10.1016/j.cie.2012.06.015>
- [19] Parisio A., Jones CN., *A two-stage stochastic programming approach to employee scheduling in retail outlets with uncertain demand*, Omega, Vol. 53, Elsevier, 2015, 97-103. <https://doi.org/10.1016/j.omega.2015.01.003>
- [20] Roeva O., Atanassova V., *Cuckoo search algorithm for model parameter identification*, Int. J. Bioautomation, Vol. 20(4), 2016, 483-492.
- [21] Soukour A., Devendeville L., Lucet C., Moukrim A., *A Memetic algorithm for staff scheduling problem in airport security service*, Expert Systems with Applications, Vol. 40(18), 2013, 7504-7512. <https://doi.org/10.1016/j.eswa.2013.06.073>

- [22] Tilahun S.L., Ngnotchouye J.M.T., *Firefly algorithm for discrete optimization problems: A survey*, Journal of Civil Engineering, Vol. 21(2), 2017, 535-545. <https://doi.org/10.1007/s12205-017-1501-1>
- [23] Toimil D., Gómes A., *Review of metaheuristics applied to heat exchanger network design*, International Transactions in Operational Research, Vol. 24(1-2), 2017, 7-26. <https://doi.org/10.1111/itor.12296>
- [24] Yang G., Tang W., Zhao R., *An uncertain workforce planning problem with job satisfaction*, Int. J. Machine Learning and Cybernetics, Springer, 2016. <https://doi.org/10.1007/s13042-016-0539-6>  
<http://rd.springer.com/article/10.1007/s13042-016-0539-6>
- [25] Zhou C., Tang W., Zhao R., *An uncertain search model for recruitment problem with enterprise performance*, J Intelligent Manufacturing, Vol. 28(3), Springer, 2014, 295-704. doi:10.1007/s10845-014-0997-1



# 7<sup>th</sup> International Workshop on Smart Energy Networks & Multi-Agent Systems

**O**UR energy supply infrastructure is in the middle of a transition from a conventional star-like energy supply topology with a manageable number of well-structured power plants towards a grid topology with a myriad of different generation units that are geographically widely distributed. Additionally, the increasing integration of volatile and intermittent renewable energy resources brings massive challenges to grid operations and its composition with respect to power system commitment, dispatching and reserve requirements.

The fact that renewable energy generation units will increase their share in the overall energy production, calls for technologies to be developed in the next decades to deal with the transition of the energy supply system and the distribution of renewable energy generation units. This includes technologies to integrate, handle and intelligently manage energy storage systems, grid load peak-shaving, smart supply system components, more efficient and intelligent coupling of heating with electrical power, heat storage, intelligent load shifting and balancing, to name only a few here.

All these have in common that the future power grid has to be intelligent, where generation and consumption units communicate or even negotiate their offer or their demand of energy through an ‘internet of energy’. Thus, to efficiently design and develop those distributed energy management systems is one of the key challenges to be solved to transform the energy supply system, addressing distributed coordination, as well as different forms of energy like electricity, heat, natural gas and other.

Information and communication technologies are the key enablers of such envisioned systems, where especially the agent-paradigm provides an excellent modelling approach for the distributed character of energy systems. Although significant efforts and investments have already been allocated into the development of smart grids, there are, however, still significant research challenges to be addressed before the promised efficiencies or visions can be realised. This includes distributed, collaborative, autonomous and intelligent software solutions for simulation, monitoring, control and optimization of smart energy networks and interactions between them.

## TOPICS

The SEN-MAS’19 Workshop aims at providing a forum for presenting and discussing recent advances and experiences in building and using multi-agent systems for modelling, simulation and management of smart energy networks. In particular, it includes (but is not limited to) the following topics of interest:

- Experiences of Smart Grid implementations by using MAS
- Applications of Smart Grid technologies
- Distributed energy management of distributed generation and storage based on MAS
- Examples of design patterns for MAS in distributed energy management systems
- Microgrids, Islands Power Systems
- Real time control of energy networks
- Distributed planning process for energy networks by using MAS
- Self-configuring or self-healing energy systems
- Load modelling and control with MAS
- Simulations of Smart Energy Networks
- Software Tools for Smart Energy Networks
- Energy Storage
- Electrical Vehicles
- Charge scheduling for electric vehicles (and fleets) based on MAS
- Interactions and exchange between networks for electricity, gas and heat
- Stability in Energy Networks
- Distributed Optimization in Energy Networks
- Safety and security issues for MAS in Smart Grids

## EVENT CHAIRS

- **Brehm, Robert**, University of Southern Denmark, Denmark
- **Derksen, Christian**, University Duisburg-Essen, Germany

## PROGRAM COMMITTEE

- **Bilal, Bilal**
- **Bremer, Joerg**, joerg.bremer@uni-oldenburg.de, Germany
- **Derksen, Christian**
- **Fortino, Giancarlo**, Università della Calabria
- **Hildmann, Hanno**, Universidad Carlos III de Madrid (UC3M), Spain
- **Karnouskos, Stamatis**, SAP, Germany
- **Klusch, Matthias**, German Research Center for Artificial Intelligence, DFKI, Germany
- **Loose, Nils**
- **Moench, Lars**, FernUniversität Hagen, Germany
- **Nieße, Astrid**, Leibniz Universität Hannover, Germany
- **Paprzycki, Marcin**, Systems Research Institute Polish Academy of Sciences, Poland

- **Redder, Mareike**
- **Sonnenschein, Michael**, Professor (retired) at the University of Oldenburg, Germany
- **Sudeikat, Jan**, Hamburg Energie GmbH, Germany
- **Vale, Zita**

# Tool-assisted Surrogate Selection for Simulation Models in Energy Systems

Stephan Balduin\*, Frauke Oest\*, Marita Blank-Babazadeh\*, Astrid Niebe<sup>‡</sup> and Sebastian Lehnhoff\*

\* OFFIS – Institute for Information Technology, Oldenburg, Germany

<sup>‡</sup> Leibniz University Hannover, Germany

Email: {frauke.oest, stephan.balduin, sebastian.lehnhoff, marita.blank-babazadeh}@offis.de  
niesse@ei.uni-hannover.de

**Abstract**—Surrogate models have proved to be a suitable replacement for complex simulation models in various applications. Runtime considerations, complexity reduction, and privacy concerns play a role in the decision to use a surrogate model. The choice of an appropriate surrogate model though is often tedious and largely dependent on the individual model properties. A tool can help to facilitate this process. To this end, we present a surrogate modeling process supporting tool that simplifies the process of generation and application of surrogate models in a co-simulation framework. We evaluate the tool in our application context, energy system co-simulation, and apply it to different simulation models from that domain with a focus on decentralized energy units.

## I. INTRODUCTION

THE simulation of smart grids is a key step in the deployment process of new technologies and methodologies in the present power system for safety and costs reasons. Co-simulation frameworks like *mosaik*<sup>1</sup> assist the simulation process in the energy domain by providing programmable interfaces for different simulation models and realizing data flow dependencies including synchronization issues. These simulation models can become quite complex and can be provided in different programming languages. This can lead to a slowed down performance of a smart grid simulation. Performance plays a role especially in large-scale setups, such as in the research projects *Smart Nord* [1] or *D-Flex* [2], which are required for the evaluation of new Smart Grid algorithms or sustainability assessments. Furthermore, simulation models might be supplied by industrial stakeholders and thus must be considered as intellectual property that should not be disclosed to partners.

A solution concerning these issues might be the use of a data-driven abstraction of simulation models, so called surrogate models. A surrogate model is a function that maps input values to one or more output values. For this purpose, machine learning algorithms can be used to determine the relation between input and output by training with sample data generated by the original simulation model [3]. The creation of those surrogate models underlies several degrees of freedom like the choice of a sampling strategy for the input data and the choice of the surrogate algorithm. The performance of a surrogate for a particular simulation model

depends not only on the specific type of that model but also can be measured differently depending on the evaluation function. An evaluation function measures the similarity of outputs between surrogate and simulation model for a set of input combinations. Based on the number of existing surrogate modeling algorithms, the identification of an appropriate surrogate can be computationally quite intensive and should be (semi-)automated to ensure replicability and comparability of the results.

For these reasons we propose to use a tool to support the selection of appropriate surrogate models. With the help of various evaluation functions, we can then evaluate the performance of different settings from specific sampling strategies and surrogate models. Furthermore, we present the Python-based open source tool *MeMoBuilder*<sup>2</sup> to support this process. *MeMoBuilder* provides a semi-automated surrogate modeling process including comparison with the original simulation model in a time series evaluation.

The rest of this paper is structured as follows. First, in Chapter II we present the surrogate modeling process and highlight the challenges that emerge from this. In Chapter III we look at existing tools and other work in the field of surrogate modeling. Chapter IV focusses on the tool *MeMoBuilder* which is our approach for reducing the complexity of the surrogate modeling process. In Chapter V we will present a case study to evaluate the tool with reflection to the defined challenges. This paper ends with conclusion and outlook on future work in Chapter VI.

## II. CHALLENGES IN SURROGATE MODELING

The process of creating a surrogate model, which is also called meta model or response surface, is well documented in the literature, e.g. in Myers et al. [4]. In the following, we briefly recap the surrogate modeling process as described by Forrester et al. [5] to point out the challenges in this process. We then derive the requirements that are important for a surrogate modeling tool.

### A. Choosing the sampling strategy

The first step of this process starts with the generation of so-called samples through sampling strategies. We use

<sup>1</sup><http://mosaik.offis.de/>

<sup>2</sup><https://github.com/stbalduin/memobuilder>

the term sampling strategy for the application of a sampling design, i.e. the theoretical construction to generate samples. These consist of a set of possible input combinations  $\mathbf{X} = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(n)}\}$  for the original simulation model  $f(\mathbf{x})$  and their corresponding outputs  $\mathbf{y}$  calculated by the simulation model. A sampling strategy defines which input combinations will be chosen to generate samples and it is important to pick the most relevant data to generate a good model. Finding an appropriate sampling strategy can depend on certain problem-specific and often contrary requirements, e.g. being deterministic, well-balanced, guaranteeing to cover the whole sampling space, or to work well with relatively few samples [6]. The goal of these strategies is to cover the relevant sample space, i.e. non-trivial (e.g. non-linear) behavior of the simulation model is included as accurate as possible. A well-balanced sampling design can be generated deterministically, but may require a large number of samples. On the other hand, a non-deterministic sampling design may work well with fewer samples, but there is no guarantee that it will cover the whole sample area. In both cases the orthogonality, i.e. the correlation of inputs, has to be considered. Simpson et al. [6] point out that the information gain of a design is balanced against the cost of experimentation, i.e. the number of samples, and lists several measures of merit which are useful to compare designs.

### B. Choosing the surrogate algorithm

In the following step, the surrogate model  $\hat{f}(\mathbf{x})$  is created by applying so called surrogate algorithms on the previously generated data. We use the term surrogate algorithm for any supervised machine learning algorithms that is capable of creating a surrogate model. A set of samples is used as training data  $(\mathbf{X}, \mathbf{y})$  for the surrogate algorithm to adjust its parameters in order to make the resulting surrogate model as similar as possible to the original simulation model. The surrogate algorithm can be picked from a large variety of machine learning algorithms with trade-offs in their characteristics, e.g. suitability for non-linear problems, suitability for high-dimensional data, complexity in the application, or in the learning phase. The latter is often partially depending on the search for optimal hyperparameters. These kind of parameters have to be set a priori to the learning process. To find the most appropriate parameter values, an exhaustive searching process with cross-validation has to be applied which means that different splits of the samples into training set and test set are evaluated. Furthermore, some surrogate algorithms use kernel functions to build the surrogate model. The choice of kernel functions and the hyperparameter tuning of these functions also have to be optimized to the given problem.

### C. Choosing the evaluation function

To evaluate the quality of the surrogate model an evaluation function is used. In general, the error  $\epsilon$  is used to describe the deviation between original simulation model and surrogate model:  $f(\mathbf{x}) = \hat{f}(\mathbf{x}) + \epsilon$ , but there are also evaluation functions that represent the quality of the model in a different way,

e.g. correlation functions. The quality of the surrogate model approximation can be evaluated by using samples as test data (which should be distinct from the training data) on the surrogate model and as well on the original simulation model. The error  $\epsilon$  can be determined by applying an evaluation function on both resulting outputs. The choice of the evaluation function has a strong influence on the ranking of surrogate models. They can be categorized into optimistic and pessimistic functions [7]. An optimistic evaluation function weights small errors more than large ones, hence might be beneficial if the error fluctuates greatly. Pessimistic evaluation function behaves the other way around, therefore they might be useful if large errors are undesirable. But there are also other characteristics, e.g. interpretability and independence of (physical) units, which should be taken into account when selecting an evaluation function for the model. It is important to know which requirements the model has to fulfill, like how critical small errors are or who will use the model afterwards, to decide which criteria should be prioritized.

### D. Requirements for tool support

Several degrees of freedom can be found in the defined process steps of surrogate model creation and evaluation, namely the choice of the sampling strategy, the choice of the surrogate algorithm, and the choice of an evaluation function used to evaluate this model. Each choice has its benefits and drawbacks. To find the most suitable combination is always depending on the given problem and cannot be generalized [5, p. 18]. For this reason, several iterations of sample generation, model creation, and model evaluation need to be performed during the surrogate modeling process until the results meet the requirements of the application context. The multi-dimensionality of the surrogate modeling process is summarized by Figure 1. For multiple but similar structured

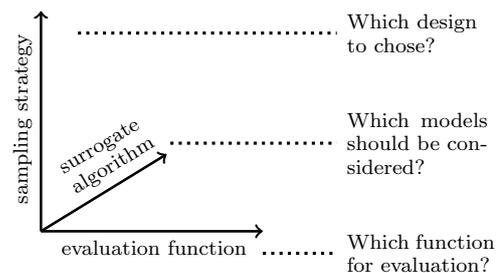


Fig. 1. Dimensions of the surrogate modeling process. Each dimension itself can be optimized quite easily, but it becomes a trade-off when all of these aspects shall be taken into account for optimization.

simulation models this is a quite repetitive process when modeling surrogates, so a tool for assistance is strongly recommended. Such a tool should fulfill the following requirements:

- R1 Support the surrogate modeling process: The tool should allow the surrogate modeler to address all degrees of freedom in experimental design, choice of surrogate algorithm, and evaluation function and thus allow for an

application specific instantiation of the surrogate modeling process.

- R2 Facilitate model exchange: The surrogate modeler is not necessarily the person who will use the model afterwards. Therefore, the tool should allow to create surrogate models which can be easily integrated into an existing environment (e.g. a co-simulation framework) and replace the original simulation models in order to easily integrate these models into smart grid simulation scenarios.
- R3 Allow modularity: In some setups, sampling data may be retrieved from other sources than available simulation models, e.g. in industry driven studies. To allow the surrogate modeler to perform only parts of the process, the tool should support a separation of concerns so that the integration of simulation models for sampling and the construction of surrogate models are independent of each other.

Although not a specific requirement in the choice or development of an appropriate surrogate modeling tool, the long-term perspective of using such a tool should be the (semi-)automatic generation of surrogate models.

### III. RELATED WORK

The whole surrogate modeling process is targeted by the Matlab Surrogate Modeling (SUMO)-Toolbox<sup>3</sup> which automates the single steps of this process. The SUMO-Toolbox builds a surrogate model of a given data source and needs only a few configurations by the user like accuracy and time constraints. However, the resulting surrogate model is bound to the Matlab environment. To deploy the model in a different setup, adaptations may be required and therefore requirement R2 is not fulfilled.

A tool aided surrogate selection is described by Mehmani et al. [8]. The authors developed the Concurrent Surrogate Model Selection (COSMOS) Framework which can be used to select an appropriate surrogate model. This tool focuses on the surrogate selection itself which comprises the optimal model type, the optimal kernel function (if needed), and optimal values of hyperparameter (if present). Despite being an important contribution in the domain of surrogate modeling, some shortcomings arise with respect to requirement R1: An easy comparison of different experimental designs is not possible.

Although not applicable to the problem of generating a surrogate for a given simulation model, the Waikato Environment for Knowledge Analysis<sup>4</sup> (Weka) proposed by Hall et al. [9] is an important open source collection of machine learning algorithms which aims to make these algorithms generally available to solve practical problems. Weka provides both a programmable and a graphical interface where no programming skills are needed when a learnable dataset is given. The focus of this tool is on data mining. Therefore, it does not contain an interface to integrate a simulation model

and automatically generate learnable data. Nonetheless, it does not support to address all degrees of freedom of the surrogate modeling process (requirement R1).

Various applications of surrogate models can be found in Koziel et al. [10], though there are no applications in the energy domain. Other works deal with the construction of a surrogate model for specific (simulation) models within concrete use cases. Pinto et al. [11] constructed a surrogate model for multi-period flexibility provided by a home energy management system. They modeled local microgeneration units, like photovoltaics, combined with flexible storage equipment which can be a battery. In their study the authors proposed an algorithm based on evolutionary particle swarm optimization to generate feasible flexibility trajectories. These trajectories were successfully used as training data for a support vector data description (SVDD) machine learning algorithm.

	SuMo Toolbox	COSMOS	Weka
R1 (Modeling process)	✓	✗	✗
R2 (Model exchange)	✗	✗	✗
R3 (Modularity)	✓	✗	(✓)

TABLE I  
SUMMARY OF THE PRESENTED TOOLS COMPARED TO OUR REQUIREMENTS.

In our research (see Table I) we did not find a tool that fulfills all requirements as defined in chapter II.

### IV. MEMOBUILDER

In this chapter we describe the architecture of the proposed Meta Model (MeMo) Builder. Our goal was to integrate surrogate algorithms, sampling strategies, and evaluation functions into one tool (requirement R1). This tool selects the optimal from each of those and generates a surrogate model that can be used within a co-simulation framework (requirement R2) as a replacement for the original simulation model. The model should also be compared with the original simulation model and behave similarly to it. For co-simulation we choose mosaik since it is a flexible tool which provides interfaces to models written in different programming languages. Thus, the surrogate modeling process is applicable regardless of the model at hand and the modeler can concentrate on the modeling process itself rather than integrating the model. The surrogate model itself can also be easily integrated in mosaik.

Mosaik facilitates a time discrete simulation, i.e. each simulation step has the same fixed length and each simulator can decide when it will be activated. Simulation models used in such a framework need to perform their simulation step for a given time interval, and a defined set of inputs and parameters as shown in Figure 2. The same applies to the surrogate models we want to build. We developed the MeMoBuilder as a prototype to identify challenges and benefits of the surrogate modeling process for simulation models in energy system, and adapt the tool to the needs identified in following this process. Further, we wanted to explore the possibilities of (semi-)automatic surrogate model generation of mosaik

<sup>3</sup><http://sumo.intec.ugent.be/>

<sup>4</sup><http://www.cs.waikato.ac.nz/ml/weka/>

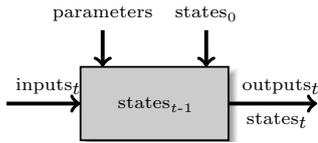


Fig. 2. The simulation model is initialized with parameters and initial states<sub>0</sub>. When simulation starts the model gets inputs<sub>t</sub> for step *t*. The results of each step are the states<sub>t</sub> which will be saved internally and then used in step *t* + 1, and outputs<sub>t</sub>.

component models for power and smart grid simulation scenarios. MeMoBuilder provides a set of sampling strategies, surrogate algorithms, and evaluation functions which can be chosen to generate a surrogate model. In Figure 3 the modular

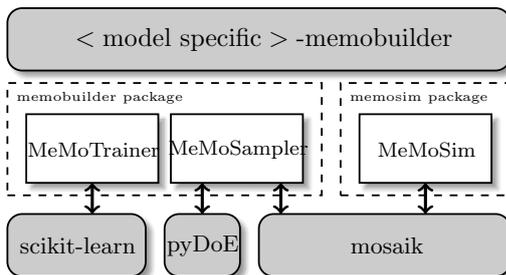


Fig. 3. MeMoBuilder has a modular structure where single components can be left out or be replaced. The core packages are the MeMoSampler and the MeMoTrainer which only need to be adapted if different frameworks for machine learning or design of experiments are used. The MeMoSim package depends on the chosen co-simulation framework which is in this case mosaik.

architecture (requirement R3) and main components of the MeMoBuilder are shown. For each simulation model, a model-specific MeMoBuilder is configured in a YAML<sup>5</sup> (YAML Ain't Markup Language) configuration file. This configuration file is also used for other degrees of freedom like how and which of the other MeMoBuilder components will be used in the surrogate generation process.

In the first step of this process the MeMoSampler uses a framework like pyDoE<sup>6</sup> to generate the sampling designs configured in the YAML file. According to these designs, the simulation model is sampled within a mosaik scenario to create one or more training sets. The use of mosaik at this point ensures that the MeMoSampler can be applied to every model with an existing mosaik adapter regardless for which simulation environment it was built originally.

Next, the training sets are used to create surrogate models. Thereby, it depends on the configuration how many surrogates will be generated. Each training set is used by the MeMoTrainer for all surrogate algorithms that are configured in the model-specific MeMoBuilder. The MeMoTrainer itself uses the scikit-learn library<sup>7</sup> [12] for model fitting and cross validation of optimal hyperparameters, but other frameworks could be integrated as well. It is also possible to use multiple

evaluation functions and in this case MeMoTrainer generates a surrogate for each function.

Ultimately, there can be a whole set of surrogates and each of these will be rated using different evaluation functions. When the surrogate model generation is finished, MeMoBuilder compares the simulation model with the surrogate model in a simple simulation scenario within the chosen co-simulation framework mosaik. The results of each simulation step are stored in a database. Additionally, a visualization of these results is generated and stored. Once the simulation is finished, it is possible to see differences in the output behavior between simulation model and surrogate model.

## V. CASE STUDY

To test the functionality of the MeMoBuilder in a practical environment and check the suitability of this tool against the requirements as defined in Section II, a selection of surrogate algorithms, sampling strategies, and evaluation functions was identified and integrated in the MeMoBuilder environment.

### A. Chosen sampling strategies

We integrated four sampling strategies of both random and deterministic type.

a) *Random based sampling strategies:* Latin Hypercube Sampling (LHS) is probably the most common strategy and has some advantages which possibly lead to the wide acceptance of this method: It is well balanced while only a small number of samples is needed [3, p. 198ff]. LHS is nearly as easy to apply as the other random-based strategy we use, the Monte Carlo Sampling (MCS) which is pure random selection of sample points. Both give the chance, but not the guarantee that the whole sample space will be covered. Furthermore, in some setups the orthogonality of these designs is not given [13, p. 42].

b) *Deterministic sampling strategies:* Besides the random based strategies, we used two deterministic strategies. Both, the sequence of Halton (HSEQ) and the sequence of Sobol (SSEQ) use prime numbers to generate a sequence of numbers, e.g. the prime number 2 generates the sequence  $\frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{5}{8}, \frac{3}{8}, \dots$ . While HSEQ varies the prime numbers to generate a new sequence, SSEQ permutes this sequence using primitive polynomials. For a more detailed explanation we refer to Lemieux [14, 157ff]. The sample space generated by deterministic designs is typically well-balanced and has only occasionally issues with missing orthogonality [13].

### B. Chosen surrogate algorithms

As surrogate algorithms we selected five heterogeneous algorithms from the field of interpolation, neural networks, and other regression methods. One of the simplest surrogate algorithm is linear regression such as LASSO which constitutes a fast polynomial approximation According to Hastie et al. [15, p. 43] this regression function can outperform more complex methods if the data is structured linearly, a small set of training data, or sparse data is used.

Another approach is to use lazy learners like the k-nearest neighbors (k-NN) algorithm. As described in Yang et al. [16]

<sup>5</sup><http://yaml.org/>

<sup>6</sup><https://pythonhosted.org/pyDOE/>

<sup>7</sup><http://scikit-learn.org/>

the  $k$  nearest neighbors of the learned samples are directly used to estimate missing outputs. According to Ertel [17, p. 199], apart from finding the correct hyperparameters,  $k$ -NN has no actual learning phase therefore it belongs to the lazy learners. For each estimated output,  $k$ -NN calculates the distance of each sample to find the  $k$  nearest samples. Ertel also points out finding the next nearest neighbors according to the given input can be computational intensive if many training samples are used. According to Samaniego and Schulz [18] its strength lies in the flexibility which makes  $k$ -NN an appropriate choice for non-linear data structures [19].

According to Cui et al. [20] Kriging is often used to interpolate data between known data points which is done by a combination of a polynomial model and a realization of a normally distributed Gaussian random process. Simpson et al. [6] state that the strength of Kriging lies in the variety of correlation functions that can be used to shape the Gaussian random process.

Support Vector Regression (SVR) is a type of support vector machines with similarities to Kriging, since the heart of both is a kernel function [21]. This algorithm uses the kernel function in order to transform non-linear regression problems into linear by mapping the original input space to a higher feature dimension space [22].

An ANN is made of several interconnected neurons which process data coming either from outside or from other neurons. The challenge in creating ANN focuses on architectural design and the number of neurons that should be used [6]. A well constructed ANN can be quite powerful in this sense that they can handle non-linear data structures [23] as well as they can handle high-dimensional data, although this can be computationally intensive. A mitigation of computing time could be achieved by parallel computing [6]. For this purpose, multi-layer perceptron regression will be used.

### C. Chosen evaluation functions

To evaluate the generated surrogate models diverse evaluation functions were selected. As a pessimistic evaluation function we choose the mean squared error (MSE) where outliers are weighted quadratically in order to punish large errors more than small errors. The mean absolute error (MAE) is punishing outliers linearly, so it is less pessimistic than the MSE. It is easier to interpret than the MSE since units are not effected by this function.

We also choose the determination function  $R^2$  which is related to the Pearson Correlation Coefficient [3, p. 113]. In contrast to error functions where the error should be minimized, in  $R^2$  a value close to 1 means a high correlation of data and therefore the surrogate is similar to the original model. Hence, an  $R^2$  close to 0 or even negative values can be interpreted as low correlation and thus the surrogate model is not well modeled. The  $R^2$  is free of units which leads to intuitive interpretations of this function. The evaluation functions described above are taken from the scikit-learn library. More information about these functions can be found in their

documentation and user guides<sup>8</sup>. In addition to the scikit-learn evaluation functions, we choose the harmonic average error (HAE), which is shown in Equation 1.

$$\text{HAE}(y, \hat{y}) = \left( \frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{\sqrt{(y_i - \hat{y}_i)^2}} \right)^{-1} \quad (1)$$

Here, the  $n$  is the number of samples,  $y$  the result of the original model, and  $\hat{y}$  the result of the surrogate model as it is the approximation of the original model. This function allows to dominate small errors over large errors which means that this metric allows to have few large errors if there are small errors to compensate. Therefore, the HAE is considered to be optimistic.

### D. Chosen simulation models

We conduct our case study using three simulation models representing different home energy system units that were already in use for different energy system simulation scenarios. These models will be briefly explained in the following without going to much into detail.

The first model is a battery which has an internal state of charge and takes the target electrical power as input. In each step, the electrical power output is calculated depending on the current state of charge. The output has a negative sign if the battery "consumes" energy, otherwise the output has positive sign.

The second model is that of a photovoltaic (PV) plant system. The model has the module temperature as internal state and uses several input variables like time stamp, solar radiation, and air temperature. Based on geo and other information stored in the model, the sun position is calculated depending on the current time stamp which is then used to compute the electrical power output depending on current radiation on the surface of the PV plant.

The last model is the fuel cell (FC) which produces power and heat at the same time. We consider electrically driven operation, i.e. the FC follows a certain power output rather than a thermal profile in thermally driven operation. This model has two inputs, three outputs, and the electrical power as internal state. The inputs are the temperature of the incoming heating water and the target electrical power for the next time interval. The outgoing temperature of the heating water, the thermal power, and the actual electrical power are regarded as outputs of the model. The actual electrical power is divided into discrete fixed electrical power stages whereas the thermal outputs can have continuous values. It should be denoted that the actual electrical power does not need to be the same as the targeted electrical power due to internal restrictions of the model.

All models were build at OFFIS and for each of them we used MeMoBuilder to generate a surrogate model for all combinations of the mentioned methods which results in  $4*5*4 = 80$  surrogate models. We used a uniform sample size of 5,000.

<sup>8</sup>[https://scikit-learn.org/stable/modules/model\\_evaluation.html#regression-metrics](https://scikit-learn.org/stable/modules/model_evaluation.html#regression-metrics)

MeMoBuilder implements methods to apply cross-validation and hyperparameter optimization on this sample size. Since it is easily interpretable, we picked the best surrogate according to the  $R^2$  score [7]. This surrogate model will be compared with the original model in a simulation scenario.

### E. Results for the battery model

The best five surrogate models for the battery are shown in Table II. Note that these are the best models according to the  $R^2$ . Using a different score for sorting may result in a different order of the models. For our battery model, the best combination consists of a Latin Hypercube sampling strategy, an artificial neural network, and the mean squared error for evaluation. But the MeMoBuilder also provides the results for the other combinations. In case of the battery model we see that support vector regression would as well be an appropriate surrogate model. Next, a comparison in a simulation setup is

Sampling Strategy	Surrogate Algorithm	Train Func.	$R^2$ Score	HAE Score	MAE Score	MSE Score
LHS	ANN	MSE	0.998	$2.67 \cdot 10^{-3}$	0.024	0.005
HSEQ	SVR	MSE	0.992	$1.5 \cdot 10^{-10}$	6.535	873.0
MCS	SVR	MSE	0.992	$3.76 \cdot 10^{-3}$	5.872	802.0
MCS	SVR	HAE	0.992	$2.78 \cdot 10^{-3}$	5.909	750.5
LHS	SVR	MSE	0.991	$2.27 \cdot 10^{-2}$	6.766	863.2

TABLE II

BEST FIVE SURROGATE MODELS OF THE BATTERY ACCORDING TO THE  $R^2$  SCORE. THE COMPARISON USING DIFFERENT SCORES REVEALS THAT SVR SCORES POORLY DESPITE A VERY GOOD  $R^2$  SCORE WHEN THE MSE IS CONSIDERED.

done. Both models are supplied with a schedule of electrical power targets. The results are plotted in Figure 4. We see for the electrical output ( $P_{el}$ ) the surrogate model is quite accurate most of the time. Only at the last part the value seems to oscillate. For the internal state of charge, the surrogate model results are accurate for the first 30 - 35 steps. At that point, the target power value is set to zero which is not correctly handled by the surrogate model. After that point the deviation of the prediction increases. At about step 150 the state of charge of the surrogate model reaches zero which may be the reason for the oscillating power output of the surrogate model.

The results show the difficulties of modeling internal states which converge to certain boundaries like minimal and maximal state of charge. Therefore, we could use the MeMoBuilder to investigate further combinations, e. g. other sampling strategies as the peripheral areas of the sample space seem to be not sufficiently covered by the Latin Hypercube strategy, but this is beyond the scope of this paper.

### F. Results for the photovoltaic plant

In Table III the best five surrogate models for the PV plant are shown. The best combination in this case is a Latin Hypercube sampling, an ANN, and the  $R^2$  score, but the same combination with a Monte Carlo sampling differs only very slightly after the decimal point. However, in this case the LHS

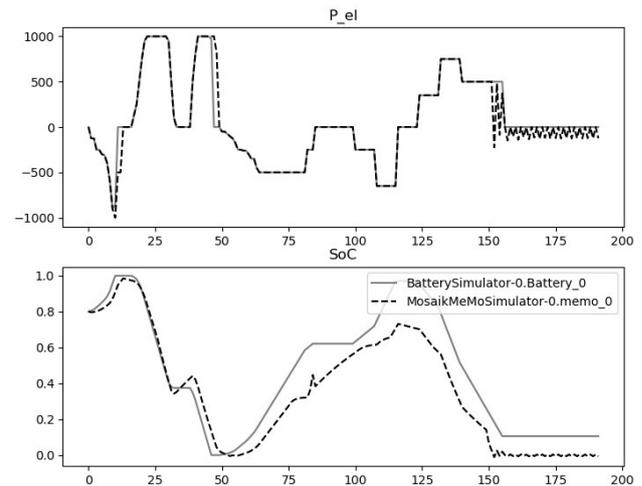


Fig. 4. Co-simulation of surrogate and original battery model comparing their electric power output and the state of charge. The grey line is the original model and the black dashed line is the surrogate model.

model is better not only at  $R^2$ , but also according to MAE and MSE scores. In the simulation, both models are provided

Sampling Strategy	Surrogate Algorithm	Train Func.	$R^2$ Score	HAE Score	MAE Score	MSE Score
LHS	ANN	$R^2$	1.0	$2.21 \cdot 10^{-2}$	0.573	1.168
MCS	ANN	$R^2$	1.0	$1.0 \cdot 10^{-10}$	1.041	3.257
MCS	ANN	MSE	0.999	$1.5 \cdot 10^{-10}$	1.108	3.111
SSEQ	ANN	MSE	0.999	$2.3 \cdot 10^{-1}$	0.639	1.533
HSEQ	ANN	$R^2$	0.999	$2.0 \cdot 10^{-10}$	0.633	1.37

TABLE III

BEST FIVE SURROGATE MODELS OF THE PV PLANT ACCORDING TO THE  $R^2$ . ANNS CONSISTENTLY DELIVER THE BEST RESULTS EVEN WHEN SORTING BY ONE OF THE OTHER SCORES.

with time stamp, radiation, and air temperature. The results are shown in Figure 5. The electrical power output prediction of the surrogate model is very accurate as long as there is actually energy generation. When there is no generation, the surrogate model predicts negative values. The module temperature seems to be captured quite accurate as well. However, the surrogate model is always one step late, but this has no visible influence on the power output.

Overall, the results for the PV plant are satisfactory in our opinion. Small flaws like the negative power output could be handled e. g. applying a  $\max(0, \hat{y})$  function on the output. Therefore, no further investigations are necessary for this model.

### G. Results for the Fuel Cell

The results of the surrogate generation for the fuel cell sorted according to the  $R^2$  score are shown in Table IV. The best combinations in this case are a Monte Carlo sampling with Kriging trained by the mean average error, and a Monte Carlo sampling with Kriging trained by the  $R^2$  score. In the

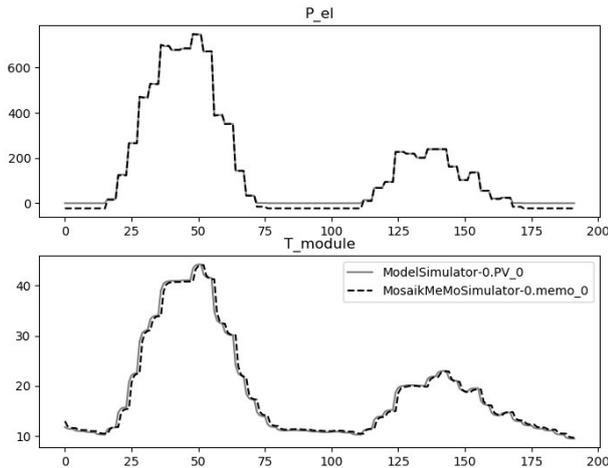


Fig. 5. Co-simulation of surrogate and original PV model comparing their electric power output. Grey line: original model, black dashed line: surrogate model.

simulation, the first configuration is used for comparison with the original model.

Sampling Strategy	Surrogate Algorithm	Train Func.	$R^2$ Score	HAE Score	MAE Score	MSE Score
MC	Kriging	MAE	0.997	$3.3 \cdot 10^{-10}$	1.598	30.72
MC	Kriging	$R^2$	0.997	$3.3 \cdot 10^{-10}$	1.598	30.72
MC	Kriging	MSE	0.996	$4.5 \cdot 10^{-10}$	2.461	43.15
LHS	Kriging	MSE	0.991	$3.5 \cdot 10^{-10}$	1.957	76.74
LHS	Kriging	$R^2$	0.991	$3.3 \cdot 10^{-10}$	1.957	76.74

TABLE IV

BEST FIVE SURROGATE TRAINING FOR THE FUEL CELL SIMULATION MODEL ACCORDING TO THE  $R^2$  SCORE. THIS RANKING SHOWS THAT KRIGING IS DELIVERING THE BEST RESULTS FOR THE  $R^2$  SCORE. .

The input schedule for electrical power is based on the standard load profile for households provided by the BDEW<sup>9</sup> and for heating water we modeled a simplified schedule for the needs of a household.

The simulation result is shown by Figure 6 for the outputs: actual electrically power ( $P_{el}$ ), thermal power ( $P_{th}$ ), and the outgoing heating water temperature ( $T_{out}$ ). The surrogate roughly follows the behavior of the original model in the output variables  $P_{th}$  and  $T_{out}$ . However, since the electrical power  $P_{el}$  is divided into discrete power stages and internally the gradient of the power is restricted so the surrogate has difficulties to reproduce the behavior especially in the transition to other power stages. The example of the fuel cell shows the difficulty of creating adequate surrogate models for models with complex internal states.

Overall, the surrogate model is satisfactory to a limited extent. If thermal power and temperature are the outputs of interest, this model performs well. For the electrical power output, however, further investigations are required.

<sup>9</sup>Bundesverband der Energie- und Wasserwirtschaft e.V

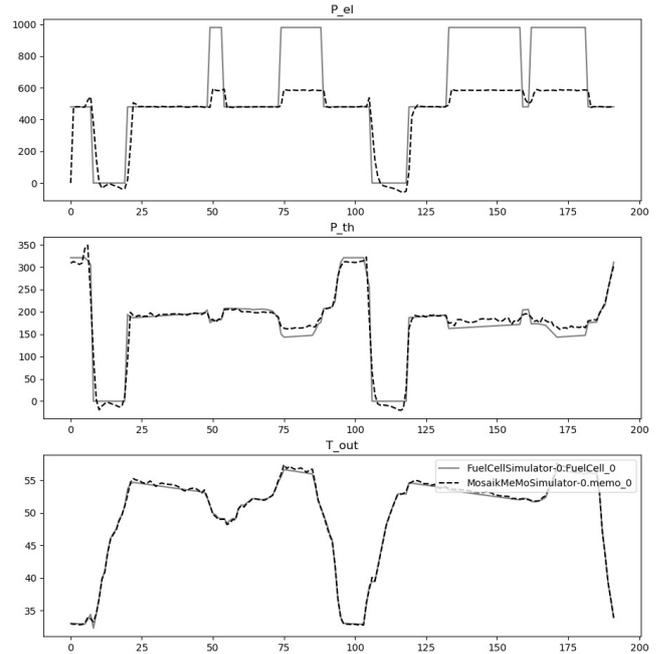


Fig. 6. Co-simulation of surrogate and original FC model. The electrical output of the model has discrete values which can be seen in the upper chart. In the middle and lower chart, the thermal outputs are shown. Grey line: original model, black dashed line: surrogate model.

## VI. CONCLUSION AND FUTURE WORK

We motivated why we need surrogate models, what the challenges of the surrogate modeling process are, and which requirements a tool has to meet in order to support this process. We presented the tool MeMoBuilder that semi-automates the surrogate modeling process while testing different combinations of sampling strategies, surrogate algorithms, and training and evaluation functions to face the challenges arising when building an appropriate surrogate model. MeMoBuilder is fully compatible to the co-simulation framework mosaik and can be used on every model which can be integrated into mosaik. An integration with other co-simulation frameworks is possible by implementing the appropriate interface between MeMoBuilder and the target framework.

Further, it is possible to separate the process of sampling and the process of training. So all the requirements as described in Section II are fulfilled. We used this tool to explore the surrogate modeling process for three different simulation models and presented the results as case study.

Our major findings are a) that specific surrogate models are more suitable for concrete simulation models than other and b) the complexity of surrogate modeling can be reduced by using a tool like the MeMoBuilder. The accuracy of the presented models range from bad to good which may depend on the choice of sorting the models according to the  $R^2$ . The MeMoBuilder provides information on which surrogate models perform well according to different criteria and gives recommendations in form of scores by different evaluation

functions.

There are still open issues which need further investigation. All provided sampling strategies and surrogate algorithms are rather generic. For some models this works quite well, for others a more specialized sampling would probably lead to better results (e.g. battery state of charge behavior). Also, the sampling designs itself are not optimized. This will be implemented in the future. Furthermore, only regression models are supported. Original models with discrete output are interpolated in the surrogate model, thus allowing values to be taken that do not exist in the original model, as shown in the fuel cell experiment. A better choice would be a classification model, but that requires training of different surrogate models for different outputs or a manual discretization of the outputs. We tried to construct the artificial neural network as a universal approximator that is generalized for many simulation models and works with a limited amount of samples. Nevertheless, there could be more suitable architectures for the individual simulation models especially with more advanced architectures like long short-term memories or convolutional neural networks.

Future studies will investigate if reducing the sample size still leads to an acceptable result since the dimensionalities of our simulation models are small. Additionally, a more advanced simulation scenario will be developed which tests the surrogate models and possible interactions with other simulation models. The next step will be integrating more specialized sampling strategies, support for classification models, and to use these surrogate models in larger scaled setups.

#### ACKNOWLEDGMENT

This work is supported by the European Community's Horizon 2020 Program (H2020/2014-2020) under project "ERIGrid" (Grant Agreement No. 654113). Further information is available at the corresponding website [www.erigrd.eu](http://www.erigrd.eu). The conception and implementation of the MeMoBuilder tool was mainly done by Thole Klingenberg.

#### REFERENCES

- [1] M. Blank, T. Breithaupt, J. Bremer, A. Dammasch, S. Garske, T. Klingenberg, S. Koch, O. Lünsdorf, A. Niesse, S. Scherfke, L. Hofmann, and M. Sonnenschein, *Smart Nord Final Report*. Uni Hannover, 4 2015, pp. 21–30.
- [2] M. Blank, M. Gandor, A. Niesse, S. Scherfke, S. Lehnhoff, and M. Sonnenschein, "Regionally-specific scenarios for smart grid simulations," in *5th International Conference on Power Engineering, Energy and Electrical Drives (POWERENG2015)*. IEEE, 5 2015, pp. 250–256. [Online]. Available: <http://dx.doi.org/10.1109/PowerEng.2015.7266328>
- [3] J. P. Kleijnen, *Design and Analysis of Simulation Experiments*. Springer International Publishing, 2015. [Online]. Available: <https://doi.org/10.1007/978-3-319-18087-8>
- [4] R. H. Myers, D. C. Montgomery, and C. M. Anderson-Cook, *Response surface methodology: process and product optimization using designed experiments*. John Wiley & Sons, 2016.
- [5] A. I. J. Forrester, A. Söbester, and A. Keane, *Engineering Design via Surrogate Modelling - A Practical Guide*. Wiley, 2008.
- [6] T. Simpson, J. Poplinski, P. N. Koch, and J. Allen, "Metamodels for computer-based engineering design: Survey and recommendations," *Engineering with Computers*, vol. 17, no. 2, pp. 129–150, jul 2001. [Online]. Available: <https://doi.org/10.1007/s00366-001-0138-1>
- [7] D. Gorissen, I. Couckuyt, E. Laermans, and T. Dhaene, "Multiobjective global surrogate modeling, dealing with the 5-percent problem," *Engineering with Computers*, vol. 26, no. 1, pp. 81–98, aug 2009. [Online]. Available: <https://doi.org/10.1007/s00366-009-0138-1>
- [8] A. Mehmami, S. Chowdhury, C. Meinrenken, and A. Messac, "Concurrent surrogate model selection (COSMOS): optimizing model type, kernel function, and hyper-parameters," *Structural and Multidisciplinary Optimization*, vol. 57, no. 3, pp. 1093–1114, sep 2017. [Online]. Available: <https://doi.org/10.1007/s00158-017-1797-y>
- [9] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016. [Online]. Available: <https://doi.org/10.1016/b978-0-12-804291-5.00024-6>
- [10] S. Koziel, S. Ogurtsov, and L. Leifsson, *Surrogate-Based Modeling and Optimization*. Springer New York, 2013. [Online]. Available: <https://doi.org/10.1007/978-1-4614-7551-4>
- [11] R. Pinto, R. J. Bessa, and M. A. Matos, "Surrogate model of multi-period flexibility from a home energy management system," *CoRR*, *abs/1703.08825*, 2017.
- [12] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011. [Online]. Available: <http://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>
- [13] K. Siebertz, D. van Bebber, and T. Hochkirchen, *Statistische Versuchsplanung - Design of Experiments (DoE)*. Springer, 2017. [Online]. Available: <https://doi.org/10.1007/978-3-662-55743-3>
- [14] C. Lemieux, *Monte carlo and quasi-monte carlo sampling*. Springer Science & Business Media, 2009. [Online]. Available: <https://doi.org/10.1007/978-0-387-78165-5>
- [15] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer-Verlag, 2009.
- [16] L. Yang, S. Liu, S. Tsoka, and L. G. Papageorgiou, "Mathematical programming for piecewise linear regression analysis," *Expert systems with applications*, vol. 44, pp. 156–167, 2016.
- [17] W. Ertel, *Grundkurs Künstliche Intelligenz - Eine praxisorientierte Einführung*. Springer Vieweg, 2013. [Online]. Available: <https://doi.org/10.1007/978-3-658-13549-2>
- [18] L. Samaniego and K. Schulz, "Supervised classification of agricultural land cover using a modified k-NN technique (MNN) and landsat remote sensing imagery," *Remote Sensing*, vol. 1, no. 4, pp. 875–895, nov 2009. [Online]. Available: <https://doi.org/10.3390/2Fr1040875>
- [19] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning*. Springer, 2013, vol. 112.
- [20] C. Cui, M. Hu, J. D. Weir, and T. Wu, "A recommendation system for meta-modeling: A meta-learning based approach," *Expert Systems with Applications*, vol. 46, pp. 33–44, mar 2016. [Online]. Available: <https://doi.org/10.1016/j.eswa.2015.10.021>
- [21] K. Markov and T. Matsui, "Music genre and emotion recognition using gaussian processes," *IEEE Access*, vol. 2, pp. 688–697, 2014. [Online]. Available: <https://doi.org/10.1109/ACCESS.2014.2333095>
- [22] C. Hultquist, G. Chen, and K. Zhao, "A comparison of gaussian process regression, random forests and support vector regression for burn severity assessment in diseased forests," *Remote Sensing Letters*, vol. 5, no. 8, pp. 723–732, aug 2014. [Online]. Available: <https://doi.org/10.1080/2F2150704x.2014.963733>
- [23] J. V. Tu, "Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes," *Journal of Clinical Epidemiology*, vol. 49, no. 11, pp. 1225–1231, nov 1996. [Online]. Available: <https://doi.org/10.1016/2Fs0895-4356%2896%2900002-9>

# Towards fully Decentralized Multi-Objective Energy Scheduling

Jörg Bremer

Department of Computing Science  
Carl von Ossietzky University  
Oldenburg, Germany  
joerg.bremer@uni-oldenburg.de

Sebastian Lehnhoff

Department of Computing Science  
Carl von Ossietzky University  
Oldenburg, Germany  
sebastian.lehnhoff@uni-oldenburg.de

**Abstract**—Future demand for managing a huge number of individually operating small and often volatile energy resources within the smart grid is preponderantly answered by involving decentralized orchestration methods for planning and scheduling. Many planning and scheduling problems are of a multi-objective nature. For the single-objective case – e.g. predictive scheduling with the goal of jointly resembling a wanted target schedule – fully decentralized algorithms with self-organizing agents exist. We extend this paradigm towards fully decentralized agent-based multi-objective scheduling for energy resources e.g. in virtual power plants for which special local constraint-handling techniques are needed. We integrate algorithmic elements from the well-known  $S$ -metric selection evolutionary multi-objective algorithm into a gossiping-based combinatorial optimization heuristic that works with agents for the single-objective case and derive a number of challenges that have to be solved for fully decentralized multi-objective optimization. We present a first solution approach based on the combinatorial optimization heuristics for agents and demonstrate viability and applicability in several simulation scenarios.

## I. INTRODUCTION

The upcoming smart grid gives rise to several multi-objective control tasks. Due to the expected huge number of distributed energy resources (DER) that have to be controlled, self-organized and decentralized algorithms are seen as the most promising solution. On the other hand, surprisingly low effort has been put in developing decentralized multi-objective approaches so far. Here, we go with the example of the predictive scheduling problem [1].

To enable small and individually operated energy devices to responsibly take over control tasks, pooling of different DER is necessary in order to gain enough potential and flexibility. An established concept for such pooling is the virtual power plant (VPP) [2], [3]. Orchestration of such groups of energy units is done by different scheduling procedures that frequently involve multi-objective optimization.

Predictive scheduling [1] describes an optimization problem for day-ahead planning of energy generation in VPPs, where the goal is to select a schedule for each energy unit – from an individual search space of feasible schedules with respect to a future planning horizon – such that as a global objective the distance to a target power profile for the VPP is minimized (e.g. a product from an energy market). Actually, this constitutes a multi-objective problem. Further objectives

like cost minimization, maximization of residual flexibility (for later planning periods) or environmental impact are usually to be achieved concurrently [4]. These goals are often conflicting.

So far, the problem is often reduced to the single objective case for proper solving; if applicable with a combination of different objectives to a single, weighted sum of objectives. We propose a fully decentralized multi-objective algorithm for this problem based on concepts from the combinatorial optimization heuristics for agents (COHDA) and the  $S$ -metric selection evolutionary multi-objective algorithm (SMS-EMOA). The goal is to derive a self-organization approach that results in autonomously acting agents that determine a Pareto front (or at least an approximation) without any central control.

The rest of the paper is organized as follows: We recap multi-objective optimization in general, a centralized solution to the predictive scheduling problem and the single objective approach to the decentralized solution. We define the set of challenges that have to be solved to make algorithms like COHDA multi-objective capable and present a first solution approach that integrates concepts from the SMS-EMOA. We conclude with evaluation results from a simulation study and deduce some further research questions.

## II. RELATED WORK

Decentralized, multi-agent-based multi-objective optimization has so far not gained much attention; at least in the sense of jointly calculating the Pareto front of a given problem. Some approaches have been developed for tuning a multi-agent system or for internal multi-objective decisions. In [5] an example is given for the emergency response planning problem, [6] presents an elevator control. These approaches use centralized algorithms for the multi-objective part. [7] gives an example for multi-objective reinforcement learning. A decision model for objective relationships after intra-agent multi-objective solving is presented in [8].

On the other hand, some approaches have been developed for acceleration by distributing fitness evaluation in multi-objective optimization. An example for a general framework is given in [9]. But, this is not decentralized problem solving by local, agent-based decisions in a collaboration scenario, as we strive for.

### A. Predictive scheduling

We here consider rather small, distributed electricity producers that are supposed to pool together with likewise distributed electricity consumers and prosumers (like batteries) in order to jointly gain more degrees of freedom in choosing load profiles. In this way, they become a controllable entity with sufficient market power. In order to manage such a pool of DER, the following distributed optimization problem has to be frequently solved: A partition of a demanded aggregate schedule has to be determined in order to fairly distribute the load among all participating DER. A schedule  $x$  is a real valued vector with each element  $x_i$  denoting the amount of energy generated or consumed during the  $i$ th time interval withing the planning horizon. Optimality usually refers to local (individual cost) as well as to global (e.g. environmental impact) objectives in addition to the main goal: Resemble the wanted overall load schedule as close as possible.

In [10], a support vector decoder has been introduced to cope with individual constraints of different types of energy units. Constraints may be technically rooted like the state of charge of attached batteries or thermal buffer stores or be economically soft rooted or be due to individual preferences. The basic idea is to learn the enclosing envelope around the set of feasible schedules in data space and to derive a formal description that allows mapping any given schedule to or into the feasible regions. In this way solution repair and space mapping can be achieved. Such constraint handling technique is in general referred to as decoder [11], [12]. Formally, a decoder function  $\gamma$  with

$$\begin{aligned} \gamma : [0, 1]^d &\rightarrow \mathcal{F}_{[0,1]} \subseteq [0, 1]^d \\ x &\mapsto \gamma(x) \end{aligned} \quad (1)$$

transforms any given (maybe in-feasible) schedule (scaled to  $[0, 1]^d$ ) into a feasible one. Thus, the scheduling problem is transformed into an unconstrained formulation when using a decoder:

$$\delta \left( \sum_{i=1}^d s_i \circ \gamma_i(x'_i), \zeta \right) \rightarrow \min, \quad (2)$$

where  $\gamma_i$  denotes the decoder of unit  $i$  that produces feasible schedules  $x'_i \in [0, 1]^d$  and  $s_i$  scales these schedules entrywise to correct power values resulting in schedules that are operable by that unit. Technically, scaling can also be integrated into the decoding process by combining both functions. Thus, for the rest of the paper we refer with  $\gamma$  to a decoder function that maps an infeasible schedule into the feasible region and scales it appropriately to the rated power of the respective energy unit. Please note that this constitutes only a single objective solution and multi-objective scenarios so far have to combine different objectives to a single one by a weighted aggregation. Unfortunately, this is not possible in case of a mixture of global and local objectives.

For the single objective case several solutions exist. In [13] an example for a centralized approach can be found, examples for decentralized approaches are given in [14]–[16]. A centralized multi-objective variant based on parallel

tempering that harnesses a decoder extension to co-encode different key performance indicators can be found in [17]. On the other hand, several approaches neglecting or relaxing individual constraint-handling can be found [2], [18].

For multi-objective optimization in general many approaches exist. In optimization problems with more than one and at least two conflicting objectives, Pareto optimization has become an appropriate means for solving [19]. As improving on one objective degrades each conflicting one, multi-objective optimization deals with finding a set of Pareto optimal solutions as trade-off between opposing solutions. Different algorithms have been designed to find an approximation to the Pareto-optimal set  $M = \{x \in \mathbb{S} \mid \nexists x^* \in \mathbb{S} : x^* \prec x\}$  for a set of objective functions  $f_{1,\dots,n} : \mathbb{S} \rightarrow \mathbb{R}$  defined on some search space  $\mathbb{S}$  [20]; and with  $x \prec x^*$  denoting that  $x$  dominates  $x^*$ , i.e. all objective values of  $x$  are better than  $x^*$ . Different algorithms have been proposed [20]; among them are evolutionary algorithms [21], [22], genetic algorithms including the famous NSGA-II [23], or swarm-based approaches [24].

Predictive scheduling imposes some special needs on constraint handling to ensure that all local schedules are within the feasible phase-space of the individual energy resources [1], [25]. For constraint-handling in multi-objective optimization two general concepts are usually applied [26]. Either a penalty [27], [28] is added to each objective function degrading constraint violating solutions or the definition of Pareto-dominance is extended to take into account constraint violation [26], [29]. Introducing a penalty term changes the objective function and as in multi-objective optimization the impact on different objectives has to be balanced, a too weak set of penalties may lead to infeasible solutions whereas a too strong impact leads to poor distributions of solutions [26].

Nevertheless, all approaches for constraint integration so far need a closed form description of constraints. Constraints are given as a set of (possibly non-linear) in-equalities and equalities as well as a box-constraint demanding all parameters being from a specific range. In the smart grid domain, often no closed form descriptions of constraints are available. Such closed form description does not exist in decentralized energy resource scheduling that includes (at least in general) arbitrary unit types [30].

A first solution approach to hybridizing multi-objective optimization and decoder was given in [31], with a centralized approach based on SMS-EMOA.

### B. SMS-EMOA

Using  $S$ -metric selection for evolutionary multi-objective algorithms has first been proposed by [22]. The  $S$ -metric is based on the hypervolume encapsulated by the set of non-dominated solutions and a reference point [21] and can thus be described as the Lebesgue measure  $\Lambda$  of the union of hypercubes defined by the reference point  $x_r$  and the set of non-dominated points  $m_i$  [19], [22]:

$$S(M) = \Lambda \left( \bigcup \{a_i \mid m \in M\} \right). \quad (3)$$

---

**Algorithm 1** Basic algorithmic scheme of the SMS-EMOA (cf. [22]).

---

```

 $P^{(0)} \leftarrow \text{randomPopulation}()$ 
 $t \leftarrow 0$ 
while  $t < \text{max iterations}$  do
   $o \leftarrow \text{mutate} \circ \text{crossover}(P^{(t)})$ 
   $\{R_1, \dots, R_k\} \leftarrow \text{fast-nondominated-sort}(P^{(t)} \cup \{o\})$ 
   $p \leftarrow \arg \min_{s \in R_k} [\Delta_S(s, R_k)]$ 
   $P^{(t+1)} \leftarrow P^{(t)} \setminus \{p\}$ 
   $t \leftarrow t + 1$ 
end while

```

---

This metric constitutes an unary quality measure by mapping a solution set to a single value in  $\mathbb{R}$ : the size of the dominated space [32]. As it is desirable to have a large  $S$ -metric value for solution sets in multi-objective optimization, [32] first used this measure in a Simulated Annealing approach and [22] developed an evolution strategy (SMS-EMOA) based on this measure. Algorithm 1 shows the basic idea of SMS-EMOA. The algorithm repeatedly evolves a population of  $\mu$  solutions. In each iteration, first a new solution is generated and added to the population. Subsequently, a selection process is started to find the worst individual in the solution which is then removed from the population. Thus, the number of individuals stays constant from a steady state perspective. Selection is done by first issuing a fast non-dominated sort after [23]. In this way, the Pareto fronts are ranked and from the front with the lowest rank the individual with the lowest contribution to the hypervolume (measured by the  $S$ -metric) is removed. This process is repeated until some stopping criterion – e. g. a number of maximum objective evaluations – is met.

In [31] the latter has already been hybridized with a decoder approach for flexibility modeling and constraint-handling in multi-objective energy management.

In general, two types of objective have to be considered. In [30] constraints have been identified on different locality levels. The same holds true for objectives in a VPP. Objectives on a global as well as on a local level have to be integrated. Objectives on a global level have to be achieved jointly. An example is given by the minimization of the deviation of the aggregated joint schedule from a given product schedule that has to be delivered as contracted. These objectives can only be achieved with joint effort. In contrast, local objectives like individual cost minimization are also to be integrated. Although evaluation can only be performed locally (individual cost), help from other to be able to choose a cheaper schedule is often necessary to achieve the goal. In the following we denote with  $f$  local and with  $F$  global objectives.

### C. COHDA

In general, decentralized algorithms are considered advantageous in many fields of smart grid computation [33], [34]. For the case of predictive scheduling, [35] developed a decentralized algorithm for constrained combinatorial problems: The Combinatorial Optimization Heuristics for Distributed

Agents (COHDA). Combined with an appropriate abstraction for individual flexibilities [10]. So far, this fully decentralized approach works with a single objective function and integrates multiple objectives only by combining different objectives into a single one as weighted sum. We extended COHDA to a full-fledged decentralized multi-objective optimization algorithm.

COHDA has been designed as a fully distributed solution to the predictive scheduling problem (as distributed constraint optimization formulation) in smart grid management [36]. Each agent in the multi-agent system is in charge of controlling exactly one distributed energy resource (generator or controllable consumer) with procurement for negotiating the energy. All energy resources are drawn together to a virtual power plant and the controlling agents form a coalition that has to control the VPP in a distributed way.

An agent in COHDA does not represent a complete solution as it is the case for instance in population-based approaches [37], [38]. Each agent represents a class within a multiple choice knapsack combinatorial problem [39]. Applied to predictive scheduling each class refers to the feasible region in the solution space of the respective energy unit. Each agent chooses schedules as solution candidate only from the set of feasible schedules that belongs to the DER controlled by this agent. This selection is done according to local constraints and to a given objective that usually reflects solely the distance (dissimilarity) of the sum of this selection and the schedules of all other agents to the given target schedule.

Each agent is connected with a rather small subset of other agents from the multi-agent system and may only communicate with agents from this limited neighborhood. The neighborhood (communication network) is defined by a small world graph [40]. As long as this graph is at least simply connected, each agent collects information from the direct neighborhood and as each received message also contains (not necessarily up-to-date) information from the transitive neighborhood, each agent may accumulate information about the choices of other agents and thus gains his own local belief of the aggregated schedule that the other agents are going to operate. With this belief each agent may choose a schedule for the own controlled energy unit in a way that the coalition is put forward best while at the same time own constraints are obeyed and own interests are pursued. Thus, we have a multi-objective optimization problem when deciding on the best schedule.

All choices for own schedules are rooted in incomplete knowledge and beliefs in what other agents do; gathered from received messages. The taken choice (together with the basis for decision-making that has been received with prior messages) is communicated to all neighboring agents and, in this way, knowledge is successively spread throughout the coalition without any central memory. Thus, COHDA is a type of gossiping algorithm [41].

Each information update results in recalculating the own best schedule contribution and spreading it to the direct neighbors. By and by all agents accumulate complete information and as soon as no agent is capable of offering a schedule

leading to a better solution, the algorithm converges and terminates. Convergence has been proved in [42].

More formally, each time an agent receives a message, three successive steps are conducted. First, during the perceive phase an agent  $a_j$  updates its own working memory  $\kappa_j$  with the received working memory  $\kappa_i$  from agent  $a_i$ . From the foreign working memory the objective of the optimization (i. e. the target schedule) is imported (if not already known) as well as the configuration that constitutes the calculation base of neighboring agent  $a_i$ . An update is conducted if the received configuration is larger or has achieved a better objective value, what is only directly possible with a single-objective. In this way, schedules that reflect the so far best choices of other agents and that are not already known in the own working memory are imported from the received memory.

During the decision phase agent  $a_j$  has to decide on the best choice for its own schedule based on the updated belief about the system state  $\mathfrak{S}_k^{(a_j)}$ . Index  $k$  indicates the age of the system state information. The agent knows from a subset of (or from all) other agents, which schedules they are going to operate (the system state  $\mathfrak{S}(a_j)_k$ ). Thus, the schedule that fills the gap to the desired target schedule exactly can be easily identified. Due to operational constraints of the controlled DER, this optimal schedule can usually not be operated. In addition, other reasons might render some schedules largely unattractive due to high cost.

Because of this reason, each agent is equipped with a so called decoder that automatically maps the identified optimal schedule to a nearby feasible schedule that is operable by the DER and thus feasible. Based on a set of feasible schedules sampled from an appropriate simulation model for flexibility prediction [43], a decoder can e. g. be based directly on this set (by linearly searching the schedule with the smallest deviation) or be built by learning a support vector model after the approach of [10]. Both approaches have individual advantages and drawbacks regarding computational complexity, search space size and accuracy. Here, we used the support vector version for efficiency reasons.

As the whole procedure is based exclusively on local decisions, each agent decides privately which schedules are taken. Private interest and preferences can be included and all information on the flexibility of the local DER is kept private.

#### D. Challenges

The COHDA approach can be adapted to many different optimization problems as has been demonstrated e. g. in [3], [30], [44], [45]. Basically, solution encoding, objective evaluation and internal, local decision method have to be adapted to the problem at hand. On the other hand, adapting to the multi-objective case entails some additional challenges that have to be solved:

*a) Solution representation:* As the goal now is a Pareto front, each agent will have to manage a set of schedules (a set of own contributions to the joint set of schedules). In the multi-objective case, determining the best own selection will

no longer work by just determining the missing difference to the target and repairing it with a decoder.

*b) Solution quality assessment:* Each time an agent in COHDA decides on a new contribution to a solution (remember: an agent represents just the local contribution, not a full solution) the quality of the solution with the old contribution is compared to the one resulting from the new contribution. This assessment is usually done using the objective function evaluating both candidates. In the multi-objective case an agent represents a set of contributions to a set of solutions, thus it is not possible to compare an old solution directly with a new one using the objective.

This problem can be overcome by using measures that evaluate the quality of a set of solutions with regard to the Pareto front. When using concepts from SMS-EMOA, the achieved hypercube volume can be used.

*c) Incomplete solutions:* During the initial setting time of COHDA, solutions are incomplete by design. COHDA has been developed for predictive scheduling. One agent starts by issuing a schedule (as local solution contribution) for the own energy resource. At this point in time, a solution consists only of this single contribution (as if it was a VPP with just a single energy resource). For the single-objective predictive scheduling case this is admissible as such a solution is always worse regarding the single objective of resembling the wanted target schedule as close as possible. After some negotiation steps, more agents join in and finally a contribution from all agents are on hand. For the multi-objective case it cannot be guaranteed that solutions with incomplete contribution are worse than complete solutions. An example may be given by minimizing cost as objective. Cost dominated by primary energy would deteriorate the objective if more energy resources joint to contribute to the solution; contradicting a minimization.

Several solutions are possible.

- 1) The protocol could be altered and each agent could be requested to calculate an initial contribution in order to avoid incomplete solutions. Depending on the problem at hand these initial contributions might be nonsense as they have been determined without knowledge on the others' decisions.
- 2) A penalty term could be added to the objective in order to deteriorate solutions based on the number of agents that still have to join. In this case the agents would need knowledge on the number of agents that are in the group.
- 3) Solutions with a larger number of contributions are always considered better regardless of the evaluation result. This might not hold for all objectives and may lead to wrong convergence directions.

For some objective functions there seems to be no issue at least if the number of agents is low enough compared with the number of network connections in between them and if the agents join in quick enough. For the simulations conducted with the first approach proposed here, after an initial deterioration of the solution quality a convergence could be observed towards

better solutions. Improved versions should integrate one of the afore mentioned approaches to guarantee convergence.

*d) Localization of objectives:* In standard COHDA each agent knows the global objective function. In multi-objective COHDA all global objective functions might also be known by all agents. On the other hand, not all objectives can be calculated by using the schedules of other agents directly, as it is the case in predictive scheduling. Calculating individual costs for example requires knowledge on private cost factors of other energy resources. Such factors are usually not known publicly nor communicated. Thus essential information for calculating the objectives is missing in a fully decentralized scenario.

In [46], an extension to the decoder approach has been proposed that is capable of annotating individual schedules with performance indicators. In [47] an ontology has been presented to capture and reliably interpret these information in a decentralized scenario. In the first approach presented here, this issue is currently neglected and only objectives that can be calculated without further information are used.

*e) Convergence detection:* In standard COHDA the solution converges to a single solution and eventually all agent represent the same solution. In the multi-objective case, all solution sets converge towards the same Pareto front and eventually all agents represent (an approximation) to the same Pareto front, but with probably different solution sets. Whether this is a problem or not highly depends on the specific problem at hand.

### III. A FIRST APPROACH

#### A. Implementation

We implemented a fully decentralized multi-objective approach for predictive scheduling by extending the COHDA algorithm. First, we had to define the solution format. A solution to the overall problem (Eq. 2) is – in the multi-objective case – a set of sets of schedules for the virtual power plant; each one consisting of a schedule for the respective energy resource in the group:

$$\mathfrak{X} = \{\mathbf{X}_1, \dots, \mathbf{X}_n\}, \quad (4)$$

with

$$\mathbf{X}_k = (x_{ij}) \in \mathbb{R}^{m \times d} \quad (5)$$

where  $x_{ij}$  denotes the mean real power of energy resource  $i$  during the  $j$ th time period. Thus each row of the matrix  $\mathbf{X}_k$  represents a schedule for the respective energy resource. This solution set  $\mathfrak{X}$  is determined in a way that it approximates the Pareto front. In this way, each agent holds and negotiates on a set of schedules for the own energy resource.

Let  $\{\mathbf{x}_1^{(a_j)}, \dots, \mathbf{x}_n^{(a_j)}\}$  denote the set of local schedules (for the own, controlled device) that is negotiated by agent  $a_j$ . Let  $\kappa_j$  be the current working memory of agent  $a_j$  (updated by an incoming message; cf. II-C). Let

$$X_O = \{\mathbf{x}_1^{(a_1)}, \dots, \mathbf{x}_n^{(a_1)}\}, \dots, \{\mathbf{x}_1^{(a_{m-1})}, \dots, \mathbf{x}_n^{(a_{m-1})}\} \in \kappa_j \quad (6)$$

be the currently known schedule selections (local solution candidates) from all the other agents  $a_1, \dots, a_{m-1} \in \mathcal{A} \setminus a_j$ . Basically, this is the system state belief  $\mathfrak{S}(a_j)_k$  without the agent's own contribution from decision  $k$ :

$$X_O = \mathfrak{S}(a_j)_k \setminus \{\mathbf{x}_1^{(a_j)}, \dots, \mathbf{x}_n^{(a_j)}\}. \quad (7)$$

Now the procedure (performed by agent  $a_j$ ) for deciding on the own schedule selection is as follows: The sum of schedules of the other agents is calculated as

$$\mathcal{O} = \{\mathbf{O}, \dots, \mathbf{O}_n\} \quad (8)$$

with

$$\mathbf{O}_i = \sum_{\mathbf{x}_i \in X_O} \mathbf{x}_i. \quad (9)$$

Now a solution of the MAS (cf. Eq. 4) to the joint Problem can be represented as

$$\left[ \begin{array}{c} \mathbf{O}_1 \\ \mathbf{x}_1^{(a_j)} \end{array} \right], \dots, \left[ \begin{array}{c} \mathbf{O}_n \\ \mathbf{x}_n^{(a_j)} \end{array} \right] \quad (10)$$

with  $\mathbf{x}_1^{(a_j)}, \dots, \mathbf{x}_n^{(a_j)}$  being the decision variables of the local problem of deciding on the best local schedules under the assumption that the other agents' schedules are operated as communicated.

Solution candidate  $\mathbf{x}_1^{(a_j)}, \dots, \mathbf{x}_n^{(a_j)}$  is initialized randomly and evolved for some iterations towards the Pareto front. For each evolution step one randomly chosen schedule  $\mathbf{x}_k \in \mathbf{x}_1^{(a_j)}, \dots, \mathbf{x}_n^{(a_j)}$  is mutated to  $\mathbf{x}'_k$  by adding a Gaussian delta  $r \in \mathcal{N}(0, 1)$  to one randomly chosen element of  $\mathbf{x}_k$ . As crossover operator, uniform crossover is applied. Please note, as an agent can only decide on its own schedules, mutation and crossover may not be applied to other agents' schedules from  $\mathbf{O}$ . Then, the agent performs a fast non-dominated sort on  $\mathbf{O} \cup \left[ \begin{array}{c} \mathbf{O}_k \\ \mathbf{x}'_k \end{array} \right]$ .

For applying the fast non-dominance-sort as introduced in [48], from the worst front the worst individual (solution with the lowest hypercube contribution) is removed. For sorting and selecting the worst individual by  $S$ -metric selection, the dominance of solutions has to be determined by using the objective functions. For this purpose, we extend the definition of dominance by integrating the decoder set:

$$\mathbf{y} \prec \mathbf{y}^* \equiv \forall i = 1, \dots, d: y_i < y_i^* \quad (11)$$

in order to keep track of the individual (technical) constraints of the energy resources by setting

$$\mathbf{y} = \left( \begin{array}{c} f_{1,1}(\gamma_1(\mathbf{x}_1)) + \dots + f_{1,n}(\gamma_n(\mathbf{x}_n)) \\ f_{2,1}(\gamma_1(\mathbf{x}_1)) + \dots + f_{2,n}(\gamma_n(\mathbf{x}_n)) \\ \vdots \\ f_{m,1}(\gamma_1(\mathbf{x}_1)) + \dots + f_{m,n}(\gamma_n(\mathbf{x}_n)) \\ F_1(\mathbf{M}) \\ \vdots \\ F_\ell(\mathbf{M}) \end{array} \right) \quad (12)$$

as introduced in [31]. In this way, variations of the previous solutions in the solution set are produced by applying variation

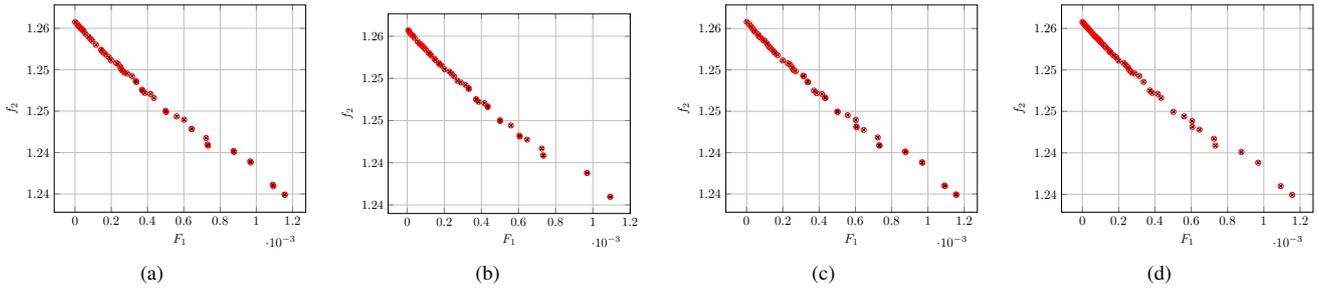


Fig. 1. Resulting solution sets of the 4 agents of the small scenario depicting the individually approximated Pareto fronts.

operators to the genotype. In this way the selection and crossover operators merely have to obey an easy to integrate box constraint that ensures that each value of a solution candidate is kept within  $[0, 1]$  if max power is scaled to 1 (corresponding to 100% rated power). No further constraints have to be integrated. Thus, the problem formulation can be regarded as constraint-free. Constraint-handling is introduced by using decoder functions that abstract from individual capabilities or technical constraints of energy units. The set of decoders ensures that selection is done on feasible solutions only and thus that the solution set approaches a Pareto front without any knowledge about controlled energy units.

The last step in the COHDA process requires comparing the result achieved in the previous round with the current achievement. If the new one is better, it is communicated to the neighboring agents, else it is discarded and the old is kept without communicating any achievement. In the single objective case, the achieved objective values can be compared directly. This is not possible in the multi-objective case as the result constitutes a set of solutions. Thus, we decided to compare both results using the hypervolume between a reference point and the solution set as rather usual in multi-objective algorithms. For fast calculation of exact hypervolumes we applied the WFG (walking fish group) algorithm [49].

With these settings we addressed all challenges identified in section II-D to constitute a fully decentralized, agent-based determination of the Pareto front of a joint multi-objective problem.

### B. Results

For our evaluation we simulated different virtual power plants consisting of different co-generation plants. The model has already been used and evaluated in different projects, e. g. [13], [46], [50], [51]. We started with a rather small setting of four agents and 96-dimensional schedules resulting in a 384-dimensional search space which has already been evaluated to be highly multi-modal and ragged [52]. As goal, two objectives were set:  $F_1$  denotes the deviation of the joint schedule from the desired target schedule ( $\|\cdot\|_2$ ) and  $f_2$  equalizes the run of the co-generation plants by minimizing peak loads:

$$f_2 = \sum_{\mathbf{x}_i \in \mathbf{X}} \sum_{j=1}^d (x_{ij} - \mu)^2, \quad (13)$$

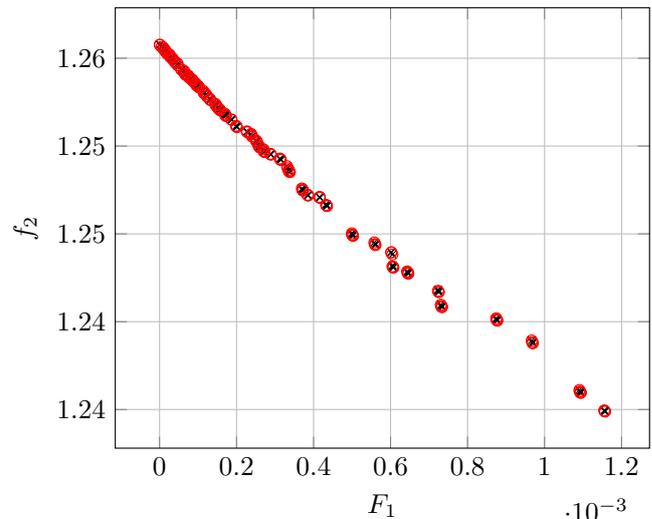


Fig. 2. Combination of local Pareto fronts of agents from solution Fig. 1. Solutions are marked with a cross; non-dominated solutions are marked by a circle. In this example, no solution is dominated by a solution from another agent.

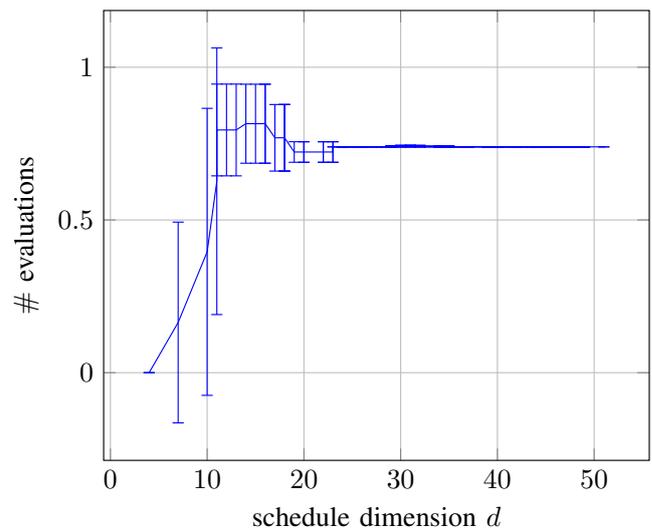


Fig. 3. Convergence and inter-agent variability (error bars) of the small scenario. Only the first 50 (out of  $\sim 1700$ ) measuring points are depicted.

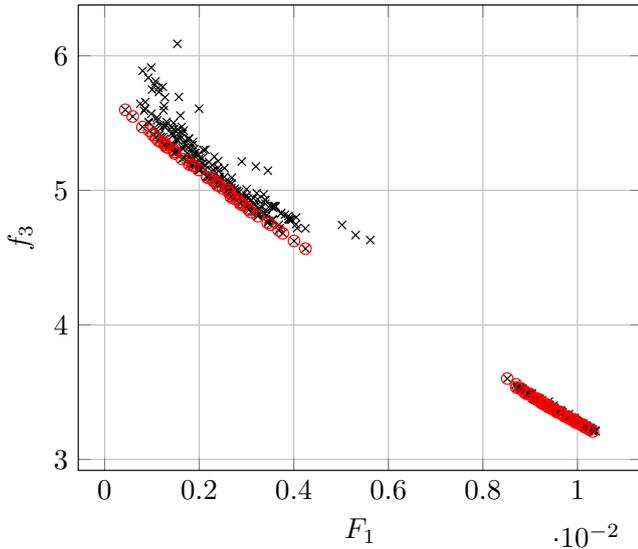


Fig. 4. Consolidated Pareto fronts from the large example. Non-dominated solutions are additionally marked with a circle. As COHDA is a heuristic, this example has obviously not completely reached a common front approximation. Maybe, a post-processing that removes dominated solutions could improve the approach significantly.

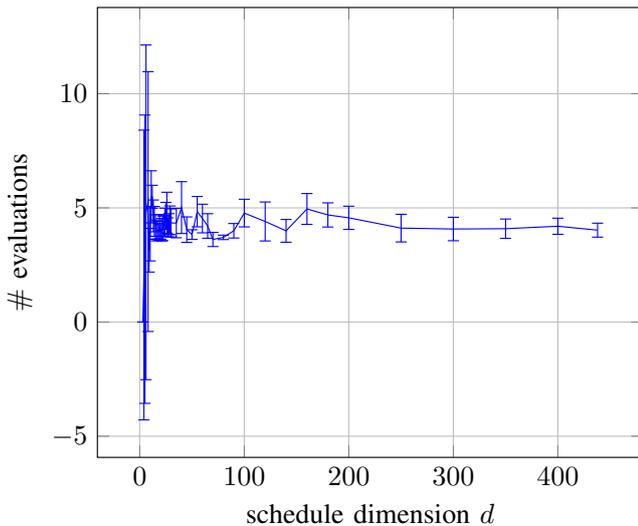


Fig. 5. Convergence and inter-agent variability (error bars) of the large scenario.

with  $\mu$  being the mean power over the whole planning horizon. In this way, the variance in power levels is minimized. Figure 1 shows an example of the resulting Pareto front approximations of the individual 4 agents. The evolutionary part during the decision phase has been run for 500 iterations each. A cross marks an individual for the solution set (50 schedules per agent) and a circle marks a non-dominated solution from the front. Note, that solution here denotes already a solution (joint schedule) for the predictive scheduling problem. Even if all schedules from the individual result sets of the agents are plotted jointly in a single plot, all solutions are non-dominated;

see Figure 2. Thus, all solutions that are individually generated by the agents stem from the same Pareto front approximation, although there are slight differences in the solution sets (regarding distribution on the front) as can be visually seen in Figure 1. The agents approximate the same front but exhibit differences in the individual solution sets.

Some statistics on this scenario with different numbers of evolutionary iterations are given in Table I.

As the proposed approach is still a heuristic, results are as yet not that perfect with growing problem size. Figure 4 shows another example from a scenario with 50 agents approximating a 50 schedule solution set each.

Here another objective has been tested: Achieve a desired state of charge (SOC) for some of the thermal buffer stores that are attached to the co-generation plants.

$$f_3 = \sum \delta(\text{SOC}_i, \text{soc}(x_i)). \quad (14)$$

In this larger scenario some of the solutions in the joint solution set are still dominated by the solutions from others, even though 5000 iterations have been conducted during decision phases. Obviously, the number of necessary iterations quickly grows with problem size: deteriorating needed negotiation time.

On the other hand, the mean hypervolume (as quality measure for a solution) converges quite early to an acceptable value. Figure 3 shows the convergences of the process that led to result 1. Here, the mean (so long achieved) hypervolume of all agents is measured at discrete points in time from the concurrently asynchronously running multi-agent system. Error bars show the inter-agents variance. At the same time the variance among all agents is determined. Depicted are only the first iterations, not the full process. Depending on the specific use case at hand, it might thus be possible to stop the negotiation at an earlier stage with a still acceptably good solution. Some more investigations will be necessary here. Figure 5 shows the situation for the second case.

TABLE I  
PERFORMANCE INDICATORS FOR BOTH TEST SCENARIOS. THIS TIME  $f_2$  WAS USED FOR BOTH.

indicator	4 agents	50 agents
hypervolume	$0.401 \pm 0.213$	$0.533 \pm 0.258$
best $F_1$	$0.103 \pm 0.110$	$0.083 \pm 0.006$
best $f_2$	$1.785 \pm 0.213$	$1.745 \pm 0.142$
# messages	$2191.6 \pm 2407.1$	$679.6 \pm 516.6$
# decisions	$1064.1 \pm 1166.6$	$328.4 \pm 245.3$

Mean achieved results for 100 runs each are depicted in Table I. The achieved hypervolume in the larger scenario larger due to the better aggregated schedule ( $F_1$ ). This observation is consistent with the single-objective case and rooted in the higher flexibility of larger VPPs. The number of exchanged messages and decisions decreases significantly with growing scenario size probably allowing for more complex decision routines of the agents in future improvements.

#### IV. CONCLUSION

For many applications within the smart grid scheduling domain, multi-objective optimization problems have to be solved. As for scalability reasons decentralized (agent-based) algorithms are seen as a promising solution, multi-objective capability has to be integrated into these methods. At the same time, proper constraint-handling is indispensable for acceptance.

With the approach demonstrated here, we integrated multi-objective capabilities taken from SMS-EMOA into fully decentralized energy scheduling.

Applicability and effectiveness of the proposed approach have been demonstrated by simulations. Nevertheless, some questions remain for further research: Replacing the decoder decision by a more complex optimization based decision method entails the need for tuning additional parameters; e. g. the (probably adaptive) number of iterations during each decision procedure. Maybe a substantial acceleration could be reached with additional convergence detection methods. Designing mutation, crossover and selection is also still subject to specialized improvements. Moreover, an integration of indicators into the communicated solution data format could improve privacy as local objectives (and thus their calculation details) would no longer be publicly known. Nevertheless, this first approach demonstrated the general feasibility of fully decentralized multi-objective optimization.

#### REFERENCES

- [1] M. Sonnenschein, O. Lünsdorf, J. Bremer, and M. Tröschel, "Decentralized control of units in smart grids for the support of renewable energy supply," *Environmental Impact Assessment Review*, no. 0, pp. –, 2014, in press.
- [2] Ł. B. Nikonowicz and J. Milewski, "Virtual power plants – general review: structure, application and optimization." *Journal of Power Technologies*, vol. 92, no. 3, 2012.
- [3] A. Nieße, S. Beer, J. Bremer, C. Hinrichs, O. Lünsdorf, and M. Sonnenschein, "Conjoint dynamic aggregation and scheduling for dynamic virtual power plants," in *Federated Conference on Computer Science and Information Systems - FedCSIS 2014, Warsaw, Poland*, M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds., 9 2014.
- [4] J. Bremer and M. Sonnenschein, "Automatic reconstruction of performance indicators from support vector based search space models in distributed real power planning scenarios," in *Informatik 2013, 43. Jahrestagung der Gesellschaft für Informatik e.V. (GI), Informatik angepasst an Mensch, Organisation und Umwelt, 16.-20. September 2013, Koblenz*, ser. LNI, M. Horbach, Ed., vol. 220. GI, 2013, pp. 1441–1454.
- [5] G. Narzisi, V. Mysore, and B. Mishra, "Multi-objective evolutionary optimization of agent-based models: An application to emergency response planning," in *Computational Intelligence*. IASTED/ACTA Press, 2006, pp. 228–232.
- [6] Y. Gu, "Multi-objective optimization of multi-agent elevator group control system based on real-time particle swarm optimization algorithm," *Engineering*, vol. 04, no. 07, pp. 368–378, 2012.
- [7] "Adaptive multi-objective reinforcement learning with hybrid exploration for traffic signal control based on cooperative multi-agent framework," *Engineering Applications of Artificial Intelligence*, vol. 29, pp. 134 – 151, 2014.
- [8] A. illah Mouaddib, M. Boussard, and M. Bouzid, "Towards a formal framework for multi-objective multi-agent planning," in *In Proc. of the 6th Int. Conf. on Autonomous Agents and Multiagent Systems*, 2007, pp. 801–808.
- [9] L. T. Bui, H. A. Abbass, and D. Essam, "Local models—an approach to distributed multi-objective optimization," *Comput. Optim. Appl.*, vol. 42, no. 1, pp. 105–139, Jan. 2009.
- [10] J. Bremer and M. Sonnenschein, "Constraint-handling for optimization with support vector surrogate models - A novel decoder approach," in *ICAART 2013 - Proceedings of the 5th International Conference on Agents and Artificial Intelligence, Volume 2, Barcelona, Spain, 15-18 February, 2013*, J. Filipe and A. L. N. Fred, Eds. SciTePress, 2013, pp. 91–100.
- [11] C. A. Coello Coello, "Theoretical and numerical constraint-handling techniques used with evolutionary algorithms: a survey of the state of the art," *Computer Methods in Applied Mechanics and Engineering*, vol. 191, no. 11-12, pp. 1245–1287, Jan. 2002.
- [12] S. Koziel and Z. Michalewicz, "Evolutionary algorithms, homomorphous mappings, and constrained parameter optimization," *Evol. Comput.*, vol. 7, pp. 19–44, 03 1999.
- [13] C. Hinrichs, J. Bremer, S. Martens, and M. Sonnenschein, "Partitioning the data domain of combinatorial problems for sequential optimization," in *FedCSIS*, M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds., 2016, pp. 551–559.
- [14] J. Bremer and M. Sonnenschein, "A distributed greedy algorithm for constraint-based scheduling of energy resources," in *Federated Conference on Computer Science and Information Systems - FedCSIS 2012, Wroclaw, Poland, 9-12 September 2012, Proceedings*, M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds., 2012, pp. 1285–1292.
- [15] A. Nieße, C. Hinrichs, J. Bremer, and M. Sonnenschein, "Local Soft Constraints in Distributed Energy Scheduling," in *5th International Workshop on Smart Energy Networks & Multi-Agent Systems, Proceedings of the 2016 Federated Conference on Computer Science and Information Systems*, M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds., Gdansk, 2016.
- [16] A. Nieße and M. Tröschel, "Controlled self-organization in smart grids," in *Proceedings of the 2016 IEEE International Symposium on Systems Engineering (ISSE)*. IEEE, 2016, pp. S. 1–6.
- [17] J. Bremer and M. Sonnenschein, "Parallel tempering for constrained many criteria optimization in dynamic virtual power plants," in *2014 IEEE Symposium on Computational Intelligence Applications in Smart Grid, CIASG 2014, Orlando, FL, USA, December 9-12, 2014*. IEEE, 2014, pp. 51–58.
- [18] C.-S. Karavas, G. Kyriakarakos, K. Arvanitis, and G. Papadakis, "A multi-agent decentralized energy management system based on distributed intelligence for the design and control of autonomous polygeneration microgrids," *Energy Conversion and Management*, vol. 103, 10 2015.
- [19] C. A. Coello Coello, G. B. Lamont, and D. A. V. Veldhuizen, *Evolutionary Algorithms for Solving Multi-Objective Problems (Genetic and Evolutionary Computation)*. Berlin, Heidelberg: Springer-Verlag, 2006.
- [20] A. Zhou, B.-Y. Qu, H. Li, S.-Z. Zhao, P. N. Suganthan, and Q. Zhang, "Multiobjective evolutionary algorithms: A survey of the state of the art," *Swarm and Evolutionary Computation*, vol. 1, no. 1, pp. 32 – 49, 2011.
- [21] E. Zitzler and L. Thiele, "Multiobjective optimization using evolutionary algorithms — a comparative case study," in *Parallel Problem Solving from Nature — PPSN V*, A. E. Eiben, T. Bäck, M. Schoenauer, and H.-P. Schwefel, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998, pp. 292–301.
- [22] M. Emmerich, N. Beume, and B. Naujoks, "An emo algorithm using the hypervolume measure as selection criterion," in *Evolutionary Multi-Criterion Optimization*, C. A. Coello Coello, A. Hernández Aguirre, and E. Zitzler, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 62–76.
- [23] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: Nsga-ii," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, Apr 2002.
- [24] J. J. Durillo, J. García-Nieto, A. J. Nebro, C. A. Coello, F. Luna, and E. Alba, "Multi-objective particle swarm optimizers: An experimental comparison," in *Proceedings of the 5th International Conference on Evolutionary Multi-Criterion Optimization*, ser. EMO '09. Berlin, Heidelberg: Springer-Verlag, 2009, pp. 495–509.
- [25] J. Bremer, B. Rapp, and M. Sonnenschein, "Support vector based encoding of distributed energy resources' feasible load spaces," in *IEEE PES Conference on Innovative Smart Grid Technologies Europe, Chalmers Lindholmen, Gothenburg, Sweden*, 2010.
- [26] F. Snyman and M. Helbig, "Solving constrained multi-objective optimization problems with evolutionary algorithms," in *Advances in Swarm Intelligence*, Y. Tan, H. Takagi, Y. Shi, and B. Niu, Eds. Cham: Springer International Publishing, 2017, pp. 57–66.
- [27] N. Srinivas and K. Deb, "Multiobjective optimization using nondominated sorting in genetic algorithms," *Evolutionary Computation*, vol. 2, no. 3, pp. 221–248, 1994.

- [28] A. Smith and D. Coit, *Handbook of Evolutionary Computation*. Department of Industrial Engineering, University of Pittsburgh, USA: Oxford University Press and IOP Publishing, 1997, ch. Penalty Functions, p. Section C5.2.
- [29] K. Deb, *Multi-Objective Optimization Using Evolutionary Algorithms*. New York, NY, USA: John Wiley & Sons, Inc., 2001.
- [30] C. Hinrichs, J. Bremer, and M. Sonnenschein, "Distributed Hybrid Constraint Handling in Large Scale Virtual Power Plants," in *IEEE PES Conference on Innovative Smart Grid Technologies Europe (ISGT Europe 2013)*. IEEE Power & Energy Society, 2013.
- [31] J. Bremer and S. Lehnhoff, "Hybridizing s-metric selection and support vector decoder for constrained multi-objective energy management," in *Hybrid Intelligent Systems*, A. M. Madureira, A. Abraham, N. Gandhi, and M. L. Varela, Eds. Cham: Springer International Publishing, 2020, pp. 249–259.
- [32] M. Fleischer, "The measure of pareto optima applications to multi-objective metaheuristics," in *Evolutionary Multi-Criterion Optimization*, C. M. Fonseca, P. J. Fleming, E. Zitzler, L. Thiele, and K. Deb, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 519–533.
- [33] A. Nieße, S. Lehnhoff, M. Tröschel, M. Uslar, C. Wissing, H. J. Apperath, and M. Sonnenschein, "Market-based self-organized provision of active power and ancillary services: An agent-based approach for smart distribution grids," in *Complexity in Engineering (COMPENG)*, 2012, June 2012, pp. 1–5.
- [34] S. D. Ramchurn, P. Vytelingum, A. Rogers, and N. R. Jennings, "Putting the 'smarts' into the smart grid: A grand challenge for artificial intelligence," *Commun. ACM*, vol. 55, no. 4, pp. 86–97, Apr. 2012.
- [35] C. Hinrichs, S. Lehnhoff, and M. Sonnenschein, "A Decentralized Heuristic for Multiple-Choice Combinatorial Optimization Problems," in *Operations Research Proceedings 2012*. Springer, 2014, pp. 297–302.
- [36] C. Hinrichs, M. Sonnenschein, and S. Lehnhoff, "Evaluation of a Self-Organizing Heuristic for Interdependent Distributed Search Spaces," in *International Conference on Agents and Artificial Intelligence (ICAART 2013)*, J. Filipe and A. L. N. Fred, Eds., vol. Volume 1 – Agents. SciTePress, 2013, pp. 25–34.
- [37] R. Poli, J. Kennedy, and T. Blackwell, "Particle swarm optimization," *Swarm Intelligence*, vol. 1, no. 1, pp. 33–57, 2007.
- [38] D. Karaboga and B. Basturk, "A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm," *Journal of Global Optimization*, vol. 39, no. 3, pp. 459–471, Nov. 2007.
- [39] T. Lust and J. Teghem, "The multiobjective multidimensional knapsack problem: a survey and a new approach," *CoRR*, vol. abs/1007.4063, 2010.
- [40] D. Watts and S. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, no. 393, pp. 440–442, 1998.
- [41] J. Liu, B. Anderson, M. Cao, and A. Morse, "Analysis of accelerated gossip algorithms," *Automatica*, vol. 49, no. 4, pp. 873–883, 4 2013.
- [42] C. Hinrichs, "Selbstorganisierte Einsatzplanung dezentraler Akteure im Smart Grid," Ph.D. dissertation, Department for Computing Science, 2014. [Online]. Available: <http://oops.uni-oldenburg.de/1960/>
- [43] J. Bremer and M. Sonnenschein, "Sampling the search space of energy resources for self-organized, agent-based planning of active power provision," in *27th International Conference on Environmental Informatics for Environmental Protection, Sustainable Development and Risk Management, EnviroInfo 2013, Hamburg, Germany, September 2-4, 2013. Proceedings*, ser. Berichte aus der Umweltinformatik, B. Page, A. G. Fleischer, J. Göbel, and V. Wohlgemuth, Eds. Shaker, 2013, pp. 214–222.
- [44] A. Nieße and M. Sonnenschein, "Using grid related cluster schedule resemblance for energy rescheduling - goals and concepts for rescheduling of clusters in decentralized energy systems," in *SMARTGREENS*, B. Donnellan, J. F. Martins, M. Helfert, and K.-H. Krempels, Eds. SciTePress, 2013, pp. 22–31.
- [45] M. Sonnenschein, C. Hinrichs, A. Nieße, and U. Vogel, "Supporting renewable power supply through distributed coordination of energy resources," in *ICT Innovations for Sustainability*, ser. Advances in Intelligent Systems and Computing, L. M. Hilty and B. Aebischer, Eds. Springer International, 2015, vol. 310, pp. 387–404.
- [46] J. Bremer and M. Sonnenschein, "Automatic reconstruction of performance indicators from support vector based search space models in distributed real power planning scenarios," in *Informatik 2013, 43. Jahrestagung der Gesellschaft für Informatik e.V. (GI), Informatik angepasst an Mensch, Organisation und Umwelt, 16.-20. September 2013, Koblenz*, ser. LNI, M. Horbach, Ed., vol. 220. GI, 2013, pp. 1441–1454.
- [47] J. Bremer, "Ontology based description of der's learned environmental performance indicators," in *Proceedings of the 1st International Conference on Smart Grids and Green IT Systems – SmartGreens 2012*, B. Donnellan, J. P. Lopes, J. Martins, and J. Filipe, Eds. Porto, Portugal: SciTePress, 04 2012, pp. 107–112.
- [48] K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan, "A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: Nsga-ii," in *Parallel Problem Solving from Nature PPSN VI*, M. Schoenauer, K. Deb, G. Rudolph, X. Yao, E. Lutton, J. J. Merelo, and H.-P. Schwefel, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2000, pp. 849–858.
- [49] R. Lyndon While, L. Bradstreet, and L. Barone, "A fast way of calculating exact hypervolumes," *IEEE Trans. Evolutionary Computation*, vol. 16, pp. 86–95, 02 2012.
- [50] J. Neugebauer, O. Kramer, and M. Sonnenschein, "Classification cascades of overlapping feature ensembles for energy time series data," in *Proceedings of the 3rd International Workshop on Data Analytics for Renewable Energy Integration (DARE'15)*. Springer, 2015.
- [51] J. Bremer and S. Lehnhoff, *Decentralized Surplus Distribution Estimation with Weighted k-Majority Voting Games*. Cham: Springer International Publishing, 2017, pp. 327–339.
- [52] A. Nieße, J. Bremer, and S. Lehnhoff, "On local minima in distributed energy scheduling," in *Position Papers of the 2017 Federated Conference on Computer Science and Information Systems, FedCSIS 2017, Prague, Czech Republic, September 3-6, 2017.*, ser. Annals of Computer Science and Information Systems, M. Ganzha, L. A. Maciaszek, and M. Paprzycki, Eds., vol. 12.



# Advances in Computer Science & Systems

**C**SS is a FedCSIS conference track aiming at integrating and creating synergy between FedCSIS events that thematically subscribe to more technical aspects of computer science and related disciplines. The CSS track 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. Technical sessions that constitute CSS are:

- CANA'19—12<sup>th</sup> Workshop on Computer Aspects of Numerical Algorithms
- C&SS'19—6<sup>th</sup> International Conference on Cryptography and Security Systems
- LTA'19—4<sup>th</sup> International Workshop on Language Technologies and Applications
- MMAP'19—12<sup>th</sup> International Symposium on Multimedia Applications and Processing
- WAPL'19 7<sup>th</sup> Workshop on Advances in Programming Languages
- WSC'19—10<sup>th</sup> Workshop on Scalable Computing



# 12<sup>th</sup> Workshop on Computer Aspects of Numerical Algorithms

**N**UMERICAL 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 coprocessors (GPU, Intel Xeon Phi, etc.)
- 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, Jaroslaw**, Maria Curie-Sklodowska University, Poland

- **Stpiczyński, Przemysław**, Maria Curie-Sklodowska University, Poland

## PROGRAM COMMITTEE

- **Amodio, Pierluigi**, Università di Bari, Italy
- **Anastassi, Zacharias**, De Montfort University, United Kingdom
- **Brugnano, Luigi**, Università di Firenze, Italy
- **Fialko, Sergiy**, Tadeusz Kościuszko Cracow University of Technology, Poland
- **Georgiev, Krassimir**, IICT - BAS, Bulgaria
- **Gravvanis, George**, Democritus University of Thrace, Greece
- **Kozielski, Stanislaw**, Silesian University of Technology, Poland
- **Lirkov, Ivan**, Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, Bulgaria
- **Luszczek, Piotr**, University of Tennessee, United States
- **Marowka, Ami**, Bar-Ilan University, Israel
- **Petcu, Dana**, West University of Timisoara, Romania
- **Sergeichuk, Vladimir**, Institute of Mathematics of NAS of Ukraine, Ukraine
- **Srinivasan, Natesan**, Indian Institute of Technology, India
- **Tudruj, Marek**, Inst. of Comp. Science Polish Academy of Sciences/Polish-Japanese Institute of Information Technology, Poland
- **Tůma, Miroslav**, Academy of Sciences of the Czech Republic, Czech Republic
- **Vazhenin, Alexander**, University of Aizu, Japan



# Parallel cache-efficient code for computing the McCaskill partition functions

Marek Palkowski, Włodzimierz Bielecki

West Pomeranian University of Technology in Szczecin

ul. Żołnierska 49, 71-210 Szczecin, Poland

Email: mpalkowski@wi.zut.edu.pl, wbielecki@wi.zut.edu.pl

**Abstract**—We present parallel tiled optimized McCaskill’s partition functions computation code. That CPU and memory intensive dynamic programming task is within computational biology. To optimize code, we use the authorial source-to-source TRACO compiler and compare obtained code performance to that generated with the state-of-the-art PLuTo compiler based on the affine transformations framework (ATF). Although PLuTo generates tiled code with outstanding locality, it fails to parallelize tiled code. A TRACO tiling strategy uses the transitive closure of a dependence graph to avoid affine function calculation. The ISL scheduler is used to parallelize tiled loop nests. An experimental study carried out on a multi-core computer demonstrates considerable speed-up of generated code for the larger number of threads.

## I. INTRODUCTION

**D**YNAMIC programming (DP) is typically applied to optimization problems in computational biology. Code implementing such computation-intensive tasks include loop nests within the polyhedral model, which allows us to apply optimization compilers to improve code performance. However, the fact that such problems are within non-serial polyadic dynamic programming (NDPD) leads to existence of non-uniform dependences in corresponding loop nests. This limits many commonly known optimization techniques such as permutation, diamond tiling [1], or index set splitting [2] to improve cache efficiency.

One of such NDPD problems is McCaskill’s algorithm, which requires computing partition functions, which used to fold an RNA secondary structure and to find the probabilities of various sub-structures. McCaskill’s recurrence is quite similar to other NPDP RNA folding algorithms such as Nussinov’s and Zucker’s ones, which are not trivial to be optimized with automatic optimization compilers.

Today, most popular techniques to automatically generate optimal parallel code are based on affine transformations. For a given loop nest statement, an affine transformation can be presented with the following relation  $[I] \rightarrow [t = C * I + c]$ , where  $I$  is the iteration vector of the statement;  $t$  is the discrete time of the execution of iteration  $I$ ;  $C * I + c$  is the affine expression. If two statement instances have the same execution time, they can be run in parallel.

To find the unknown matrix  $C$  and unknown vector  $c$ , for each loop nest statement, on the basis of dependence relations time-partition constraints are created and resolved for elements of matrix  $C$  and elements of vector  $c$ .

State-of-the-art automatic optimizing compilers, such as PLuTo [3], have provided empirical confirmation of the success of polyhedral-based code generation and optimization. PLuTo optimizing compiler is based on the affine transformation framework (ATF), which has demonstrated considerable success in generating high-performance parallel code in particular for stencils.

ATF is also used to generate tiled code. Loop tiling for improving locality groups loop statement instances into smaller blocks (tiles) allowing reuse when the block fits in local memory. In parallel tiled code, tiles are considered as indivisible macro statements. This coarsens the granularity of parallel applications that often leads to improving the performance of an application running in parallel computers with shared memory.

ATF is applied in other compilers such as Apollo and PPCG as well as commercial R-STREAM and IBM-XL. ATF has some drawbacks, papers [4], [5], [6] present its limitations for generation of parallel cache-efficient code for bioinformatics NPDP tasks. Although PLuTo generates outstanding cache-efficient code for McCaskill’s algorithm, it is not able to generate any parallel code.

Wonnacott et al. introduced serial 3-D tiling of “mostly-tileable” loop nests of Nussinov’s RNA secondary structure prediction in paper [5] to overcome some ATF limitations. But they did not present how to parallelize code generated with a proposed technique.

Mullapudi and Bondhugula [6] have also explored automatic techniques for tiling codes that lie outside the domain of standard tiling techniques. 3-D iterative tiling for dynamic scheduling is calculated by means of reorderable reduction chains to eliminate cycles between tiles for Nussinov’s algorithm. Until now, we do not have a precise characterization of the relative domains of those techniques and it is not clear how they can be applied to parallelize McCaskill’s algorithm where target arrays are not a result of reorderable functions such as minimum or maximum.

Paper [7] presents a manual implementation of parallel McCaskill’s algorithm, but the approach is limited only to message passing architectures and does not consider locality improvement for modern multi-core machines.

Li et al. show how to use array transposition to enable better caching for Nussinov’s algorithm [8] with replacing the array reading column order to the row order. The disadvantage of

this approach is the cost of additional memory management, which is overcome by tiling strategies [9]. In paper [10], Li's method was improved, but it allows for generating only serial code.

In this paper, we introduce an alternative approach as a combination of the tile correction algorithm [11] and the ISL scheduler [12] to parallelize tiled loop nests of the McCaskill's algorithm to overcome limitations of the mentioned techniques. This approach is implemented in the TRACO compiler [13].

TRACO does not find and use any affine function to transform the loop nest. It is based on the iteration space slicing framework [14] and applies the transitive closure of a dependence graph to carry out corrections of original rectangular tiles so that all dependences available in the original loop nest are preserved under the lexicographic order of target tiles. The inter-tile dependence graph does not contain any cycle and any technique of loop nest parallelization can be used [11] to generate parallel code. We apply the commonly known loop skewing technique and use the ISL library to implement it and generate parallel tiled code implementing McCaskill's algorithm. We observe high performance and scalability of that code executed on multi-core processors. We compare obtained code performance with that of code generated with PLuTo.

## II. MCCASKILL'S ALGORITHM FOR THE PARTITION FUNCTION COMPUTATION

John S. McCaskill proposed an efficient dynamic programming algorithm to compute the partition function  $Z = \sum_P \exp(-E(P)/RT)$  over all possible nested structures  $P$  that can be formed by a given RNA sequence  $S$  with  $E(P) =$  energy of structure  $P$ ,  $R =$  gas constant, and  $T =$  temperature [15].

In this paper, we study a simplified version of the approach using a Nussinov-like energy scoring scheme, i.e., each base pair of a structure contributes a fixed energy term  $E_{bp}$  independent of its context. Given such an assumption, two dynamic programming tables  $Q$  and  $Q_{bp}$  are populated. The partition function for a sub-sequence from position  $i$  to position  $j$  is provided by  $Q_{i,j}$ . Array  $Q_{bp}$  holds the partition function of the sub-sequences, which form a base pair or 0 if base pairing is not possible.

The following recursions are used to compute the partition functions  $Q$  and  $Q_{bp}$ .

$$Q_{i,j} = Q_{i,j-1} + \sum_{i \leq k < (j-1)} Q_{i,k-1} \cdot Q_{k,j}^{bp},$$

$$Q_{i,j}^{bp} = \begin{cases} Q_{i+1,j-1} \cdot \exp(-E_{bp}/RT) & \text{if } S_i, S_j \text{ can form} \\ & \text{base pair} \\ 0 & \text{otherwise} \end{cases}.$$

Listing 1 presents the implementation of computing  $Q$  and  $Q_{bp}$ . The input data are RNA sequence  $S$  as a chain of nucleotides from the alphabet AUGC (adenine, uracil,

Listing 1. Serial loop nest implementing the McCaskill partition function computation.

---

```

if (N>=1 && l>=0 && l<=5)
  for (i=N-1; i>=0; i--)
    for (j=i+1; j<N; j++){
      Q[i][j] = Q[i][j-1];
      for (k=0; k<j-i-1; k++){
        Qbp[k+i][j] = Q[k+i+1][j-1] * ↵
          ↵ ERT * paired(k+i, j-1);
        Q[i][j] += Q[i][k+i] * ↵
          ↵ Qbp[k+i][j];
      }
    }

```

---

guanine, cytosine), minimal loop length  $l$  (i.e. minimal number of enclosed positions), energy weight of base pair  $E_{bp}$  and normalized temperature  $RT$ . The memory complexity of the arrays is  $\mathcal{O}(n^2)$ , while the time complexity of a direct implementation of this algorithm is  $\mathcal{O}(n^3)$  in the sequence of length  $N$ .

Given these partition function terms, we can find base pair probabilities as well as probabilities that a certain sub-sequence is unpaired, in the manner discussed in [16].

## III. AUTOMATIC CODE OPTIMIZATION

The code presented in Listing 1 was optimized (tiled and parallelized) by means of the TRACO compiler [13]. To tile a loop nest, TRACO forms original rectangular tiles whose size is provided by the user. Then TRACO extracts dependences available in the loop nest applying the Petit tool [17], which returns 19 dependence relations describing all the dependences in the loop nest implementing McCaskill's algorithm. Extracted dependence relations are a mathematical representation of the dependence graph whose nodes are statement instances of the loop nest, while each edge states for a dependence between a pair of nodes.

Using the union of obtained dependence relations, TRACO calculates the transitive dependence of the dependence graph, for this purpose, it uses a function implementing the algorithm presented in paper [18]. The transitive closure of a given graph,  $G$ , is a graph,  $G'$ , such that  $(i, j)$  is an edge in  $G'$  if there exists a directed path from  $i$  to  $j$  in  $G$ . It is worth noting that in general, the dependence graph is parametric – the number of nodes depends on the upper bounds of loop iterators, which are represented with parameters. So, a special algorithm should be applied to calculate the transitive closure of a parametric dependence graph.

TRACO carries out the following calculations according to the algorithm presented in paper [11]. First, applying the transitive closure of the dependence graph, it checks whether the original (rectangular) tiles are valid. A valid tile with identifier  $I$  does not contain any statement instance that is the destination of the dependence whose source belongs to

TABLE I  
EXECUTION TIME (IN SECONDS) OF THE ORIGINAL, TRACO AND PLuTo TILED CODES.

N	1 Thread			2 Threads	4 Threads	8 Threads	16 Threads	32 Threads	48 Threads
	Orig.	PLuTo	TRACO						
1000	1.6883	0.6096	0.9893	1.0862	0.4055	0.3203	0.1978	0.1500	0.1744
2000	23.1211	5.4271	11.8613	9.9558	6.4051	3.6992	2.1890	1.4105	1.2818
3000	138.5503	17.843	67.6431	48.1494	27.0968	14.9998	9.0412	5.1811	4.7234
4000	391.8773	51.9886	253.2109	169.3396	94.9911	47.8198	27.1342	14.7676	14.0896
5000	874.2910	132.3715	545.7719	378.3082	210.9896	110.7063	54.2998	34.1555	32.9895

the tile whose identifier is lexicographically greater than  $I$ . If all original tiles are valid, it directly generates target code, otherwise, it corrects original tiles so that all target tiles are valid under lexicographical order. Such a correction is realized by means of transitive closure.

It is well-known that if all tiles are valid, then the corresponding inter-tile dependence graph describing dependences among tiles is acyclic, so there exists a schedule that assigns a discrete time to the corresponding tile to execute it [19]. If two or more tiles have the same schedule time, they can be run in parallel.

To extract a valid tile schedule, we need a dependence relation, which describes inter-tile dependences. Using obtained valid tiles, TRACO forms such a relation according to the way described in paper [11]. Then TRACO finds a valid tile schedule applying the ISL scheduler [12], which uses the PLuTo scheduler with Feautrier’s one [19], [20] as fallback. The PLuTo scheduler constructs a set of independent affine schedule functions that guarantee a small dependence distance over the schedule constraints. The basic idea of Feautrier’s scheduler implemented in the ISL library is to carry as many dependences as possible in each level of a multi-dimensional schedule.

For the examined loop nest, the ISL scheduler returns the following tile schedule for each statement:  $[ii, jj, kk] \rightarrow [ii + jj]$ , which means that the tile represented with identifier  $[ii, jj, kk]$  is mapped to execution time  $[ii + jj]$ . Such a schedule corresponds to the well-known loop skewing transformation [21]. It is a convenient method to implement the wavefront method of executing a loop nest in parallel, which creates a “wave” that passes through the iteration space. Skewing changes the iteration vectors for each iteration by adding the outer loop index value to the inner one.

To generate parallel code on the tile level, to each loop statement, we apply the skewing transformation  $(ii + jj)$  to form the following schedule allowing for parallel code generation.

$$SCHED\_PAR := N \rightarrow \{ (i, j, k) \rightarrow (ii + jj, jj, kk, i, j, k) \mid constraints \},$$

where *constraints* are the constraints of a set representing target tiles for a given loop nest statement.

That schedule maps each instance of a statement to a time partition whose all tiles can be executed in parallel. TRACO passes those schedules to the input of the ISL code generator, which generates target pseudo-code. The TRACO

post-processor generates target parallel compilable code in the OpenMP standard [22], which is presented at the repository [https://github.com/markpal/hpc\\_mea](https://github.com/markpal/hpc_mea). In that code, the first loop is serial, it enumerates time partitions including target tiles. The second loop is parallel, it runs tiles belonging to a given time partition, the reminding loops are serial. Intra-tile dependences (dependences within a tile) are honored because within each target tile, statement instances are executed in lexicographical order (serially).

It is worth noting that TRACO code is less regular than that generated with PLuTo because target tiles generated with TRACO are irregular while PLuTo generates regular tiles except from boundary ones.

#### IV. EXPERIMENTAL STUDY

This section presents the results of the comparison of the performance of TRACO and PLuTo tiled codes implementing McCaskill’s algorithm. To carry out experiments, we have used a computer with the following features: Intel Xeon CPU E5-2699 v2, 3.6GHz, 24 cores, 48 Threads, 45 MB Cache, 16 GB RAM. Programs were compiled with the Intel C Compiler (icc 15.0.2) and optimized at the  $-O3$  level (more aggressive optimization recommended for loops involving intensive floating point calculations). Parallelism of target code is represented in the OpenMP standard. We discovered empirically that the best tile size for TRACO code is  $[1 \times 128 \times 16]$ , i.e., the first loop should not be tiled. For tiled code generated with PLuTo, we empirically discover that the best tile size is  $[16 \times 16 \times 16]$ .

The McCaskill loop nest can be tiled by both PLuTo and TRACO, however, only TRACO allows us to parallelize target code. Although the serial code produced with PLuTo is very cache-efficient, the compiler is unable to find any affine schedule allowing for parallel execution of generated tiles. TRACO generates valid tiles applying the transitive closure of the dependence graph built for the McCaskill loop nest, then it forms a relation, which represents inter-tile dependences. Finally, using that relation, it applies the ISL scheduler to get a valid tile schedule to generate parallel code on the tile level.

Table 1 presents execution times (in seconds) for various RNA sequence lengths. Figure 2 depicts the speed-up (a ratio of  $T_1$  over  $T_n$ , elapsed times of 1 and  $n$  threads) of tiled programs for  $N = 5000$  (roughly the size of the longest human mRNA). Analyzing the obtained results, we may conclude that the TRACO code performance overcomes that of the PLuTo serial one for eight and more threads. The worse performance

of the TRACO code for the few number of threads is caused with target code irregularity (see the previous section). The lack of parallelism limits speed-up and scalability of the PLuTo loop nest implementing McCaskill's algorithm on the modern multi-core machine used for experiments.

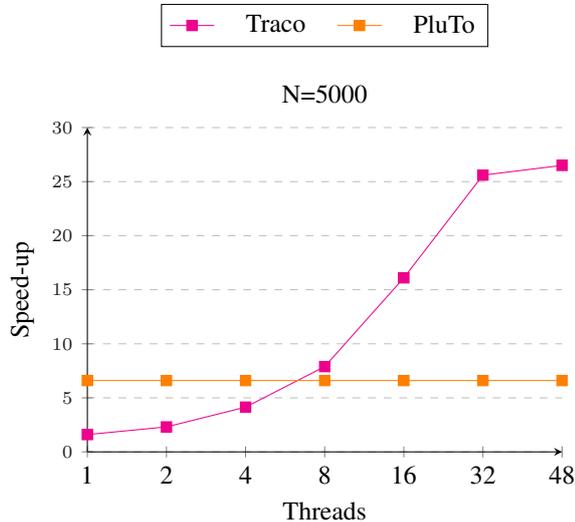


Fig. 1. Speed-up of TRACO and PLuTo codes.

## V. CONCLUSION

In this paper, we presented the usage of the TRACO compiler to optimize the loop nest implementing the McCaskill pattern function calculation. TRACO applies the transitive closure of dependence graphs to generate valid tiles under lexicographical order. Then it forms a relation describing inter-tile dependences and uses the ISL scheduler to obtain a valid tile schedule allowing for generation of parallel tiled code.

Applying optimization techniques based on affine transformations implemented in the PLuTo compiler allows for generation of only serial highly cache efficient code without any parallelism code. The proposed approach outperforms code generated with the PLuTo compiler starting up from eight threads.

It is an ongoing task to find cache efficient optimization for NPDP problems in bioinformatics with  $\mathcal{O}(n^3)$  and  $\mathcal{O}(n^4)$  complexity and non-trivial dependence patterns. In future, we plan to optimize programs implementing base pair probabilities calculation as well as prediction of their structure with maximum expected accuracy (MEA) for a given RNA sequence.

## REFERENCES

- [1] U. Bondhugula, V. Bandishti, and I. Pananilath, "Diamond tiling: Tiling techniques to maximize parallelism for stencil computations," *IEEE Transactions on Parallel and Distributed Systems*, vol. 28, no. 5, pp. 1285–1298, May 2017. doi: 10.1109/tpds.2016.2615094
- [2] U. Bondhugula, A. Acharya, and A. Cohen, "The pluto+ algorithm: A practical approach for parallelization and locality optimization of affine loop nests," *ACM Trans. Program. Lang. Syst.*, vol. 38, no. 3, pp. 12:1–12:32, Apr. 2016. doi: 10.1145/2896389
- [3] U. Bondhugula *et al.*, "A practical automatic polyhedral parallelizer and locality optimizer," *SIGPLAN Not.*, vol. 43, no. 6, pp. 101–113, Jun. 2008. doi: 10.1145/1379022.1375595 <http://pluto-compiler.sourceforge.net/>.
- [4] M. Palkowski and W. Bielecki, "Parallel tiled Nussinov RNA folding loop nest generated using both dependence graph transitive closure and loop skewing," *BMC Bioinformatics*, vol. 18, no. 1, p. 290, 2017. doi: 10.1186/s12859-017-1707-8
- [5] D. Wonnacott, T. Jin, and A. Lake, "Automatic tiling of "mostly-tileable" loop nests," in *5th International Workshop on Polyhedral Compilation Techniques*, Amsterdam, 2015.
- [6] R. T. Mullapudi and U. Bondhugula, "Tiling for dynamic scheduling," in *Proceedings of the 4th International Workshop on Polyhedral Compilation Techniques*, Vienna, Austria, Jan. 2014.
- [7] M. Fekete, I. L. Hofacker, and P. F. Stadler, "Prediction of rna base pairing probabilities on massively parallel computers," *Journal of Computational Biology*, vol. 7, no. 1-2, pp. 171–182, 2000. doi: 10.1089/10665270050081441
- [8] J. Li, S. Ranka, and S. Sahni, "Multicore and GPU algorithms for Nussinov RNA folding," *BMC Bioinformatics*, vol. 15, no. 8, p. S1, 2014. doi: 10.1186/1471-2105-15-S8-S1 [Online]. Available: <http://dx.doi.org/10.1186/1471-2105-15-S8-S1>
- [9] M. Palkowski and W. Bielecki, "Tuning iteration space slicing based tiled multi-core code implementing Nussinov's rna folding," *BMC Bioinformatics*, vol. 19, no. 1, p. 12, Jan 2018.
- [10] C. Zhao and S. Sahni, "Cache and energy efficient algorithms for nussinov's rna folding," *BMC Bioinformatics*, vol. 18, no. 15, p. 518, Dec 2017.
- [11] W. Bielecki and M. Palkowski, "Tiling of arbitrarily nested loops by means of the transitive closure of dependence graphs," *International Journal of Applied Mathematics and Computer Science (AMCS)*, vol. Vol. 26, no. 4, p. 919–939, December 2016. doi: 10.1515/amcs-2016-0065
- [12] S. Verdoolaege, "Integer set library - manual," Tech. Rep., 2011. [Online]. Available: [www.kotnet.org/~skimof/isl/manual.pdf](http://www.kotnet.org/~skimof/isl/manual.pdf).
- [13] W. Bielecki and M. Palkowski, "A parallelizing and optimizing compiler - traco," 2013. [Online]. Available: <http://traco.sourceforge.net>
- [14] W. Pugh and D. Wonnacott, "An exact method for analysis of value-based array data dependences," in *Sixth Annual Workshop on Programming Languages and Compilers for Parallel Computing*. Springer-Verlag, 1993.
- [15] M. Raden, S. M. Ali, O. S. Alkhnabshi, A. Busch, F. Costa, J. A. Davis, F. Eggenhofer, R. Gelhausen, J. Georg, S. Heyne, M. Hiller, K. Kundu, R. Kleinkauf, S. C. Lott, M. M. Mohamed, A. Mattheis, M. Miladi, A. S. Richter, S. Will, J. Wolff, P. R. Wright, and R. Backofen, "Freiburg RNA tools: a central online resource for RNA-focused research and teaching," *Nucleic Acids Research*, vol. 46, no. W1, pp. W25–W29, 2018. doi: 10.1093/nar/gky329
- [16] J. S. McCaskill, "The equilibrium partition function and base pair binding probabilities for rna secondary structure," *Biopolymers*, vol. 29, no. 6-7, pp. 1105–1119. doi: 10.1002/bip.360290621. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/bip.360290621>
- [17] W. Kelly, V. Maslov, W. Pugh, E. Rosser, T. Shpeisman, and D. Wonnacott, *New User Interface for Petit and Other Extensions*, 1996.
- [18] W. Bielecki, K. Kraska, and T. Klimek, "Using basis dependence distance vectors in the modified Floyd–Warshall algorithm," *Journal of Combinatorial Optimization*, vol. 30, no. 2, pp. 253–275, 2014.
- [19] P. Feautrier, "Some efficient solutions to the affine scheduling problem: I. one-dimensional time," *Int. J. Parallel Program.*, vol. 21, no. 5, pp. 313–348, 1992.
- [20] —, "Some efficient solutions to the affine scheduling problem: II. multidimensional time," *Int. J. Parallel Program.*, vol. 21, no. 5, pp. 389–420, 1992.
- [21] M. Wolfe, "Loops skewing: The wavefront method revisited," *International Journal of Parallel Programming*, vol. 15, no. 4, pp. 279–293, 1986.
- [22] OpenMP Architecture Review Board, "OpenMP application program interface version 4.0," 2012. [Online]. Available: [http://www.openmp.org/mp-documents/OpenMP4.0RC1\\_final.pdf](http://www.openmp.org/mp-documents/OpenMP4.0RC1_final.pdf)

# 6<sup>th</sup> International Conference on Cryptography and Security Systems

**C**RYPTOGRAPHY and security systems are two fields of security research that strongly interact and complement each other. The International Conference on Cryptography and Security Systems (CSS) is a forum of presentation of theoretical, applied research papers, case studies, implementation experiences as well as work-in-progress results in these two disciplines.

## TOPICS

The main topics of interests include:

- network security
- cryptography and data protection
- peer-to-peer security
- security of wireless sensor networks
- security of cyber physical systems
- security of Internet of Things solutions
- heterogeneous networks security
- privacy-enhancing methods
- covert channels
- steganography and watermarking for security applications
- cryptographic protocols
- security as quality of service, quality of protection
- data and application security, software security
- security models, evaluation, and verification
- formal methods in security
- trust and reputation models
- reputation systems for security applications
- intrusion tolerance
- system surveillance and enhanced security
- cybercrime: threats and countermeasures
- 5G Security
- DDoS attacks: detection and mitigation
- Security of Smart Grid systems

## EVENT CHAIRS

- **Kotulski, Zbigniew**, Warsaw University of Technology, Faculty of Electronics and Information Technology, Institute of Telecommunications, Department of Cybersecurity, Poland
- **Ksiezopolski, Bogdan**, Maria Curie-Skłodowska University, Faculty of Mathematics, Physics and Computer Science, Institute of Computer Science, Department of Cybersecurity and Polish-Japanese Academy of Information Technology, Poland

## PROGRAM COMMITTEE

- **Cabaj, Krzysztof**, Institute of Computer Science, Warsaw University of Technology, Poland

- **Caviglione, Luca**, National Research Council (CNR), Italy
- **Cheng, Shin-Ming**, National Taiwan University of Science and Technology, Taiwan
- **Courtois, Nicolas T.**, University College London, United Kingdom
- **Domingos, Maria Dulce Pedroso**, Universidade de Lisboa, Portugal
- **El Fray, Imed**, Warsaw University of Life Sciences, Faculty of Applied Informatics and Mathematics, Poland
- **Gajewski, Piotr**, Military University of Technology, Poland
- **Górski, Janusz**, Gdańsk University of Technology, Poland
- **Grochowska-Czuryło, Anna**, Poznan University of Technology, Poland
- **Gutierrez, Jaime**, Universidad de Cantabria, Spain
- **Hyla, Tomasz**, West Pomeranian University of Technology, Poland
- **Kotenko, Igor**, St.Petersburg Institute for Informatics and Automation, Russia
- **Kula, Mieczysław**, University of Silesia, Poland
- **Mauw, Sjouke**, University of Luxembourg, Luxembourg
- **Mazurczyk, Wojciech**, Warsaw University of Technology, Poland
- **Memmi, Gérard**, Telecom ParisTech, France
- **Nielek, Radosław**, Polish-Japanese Academy of Information Technology
- **Pejaś, Jerzy**, West Pomeranian University of Technology, Poland
- **Pieprzyk, Josef**, Queensland University of Technology, Australia
- **Piotrowski, Zbigniew**, Military University of Technology, Poland
- **Respício, Anna**, Universidade de Lisboa, Portugal
- **Ryan, Peter Y A**, University of Luxembourg, Luxembourg
- **Seredyński, Franciszek**, Cardinal Wyszyński University in Warsaw, Poland
- **Szałachowski, Paweł**, SUTD, Singapore
- **Tiplea, Ferucio**, Alexandru Ioan Cuza University of Iasi, Romania
- **Ustimenko, Vasyl**, Marie Curie-Skłodowska University, Poland
- **Wydra, Michał**, Lublin University of Technology, Poland



# The Low-Area FPGA Design for the Post-Quantum Cryptography Proposal Round5

Michał Andrzejczak

Military University of Technology in Warsaw  
ul. gen. Sylwestra Kaliskiego 2, 00-908 Warszawa, Poland  
Email: [michal.andrzejczak@wat.edu.pl](mailto:michal.andrzejczak@wat.edu.pl)

**Abstract**—Post-Quantum Cryptography (PQC) is getting attention recently. The main reason of this situation is the announcement by the U.S. National Institute for Standard and Technology (NIST) about an opening of the standardization process for PQC. Recently NIST published a list of submissions qualified to the second round of this process. One of the selected algorithms is Round5, offering a key encapsulation mechanism (KEM) and public key encryption (PKE). Due to high complexity of post-quantum cryptosystems, only a few FPGA implementations have been reported to date. In this paper, we report results for low-area purely-hardware implementation of Round5 targeting low-cost FPGAs.

## I. INTRODUCTION

POST-Quantum Cryptography (PQC) is an answer to a threat coming from a full-scale quantum computer able to execute Shor’s algorithm [1]. With this algorithm executed on a quantum computer, currently used public key schemes, such as RSA [2] and elliptic curve cryptosystems, are no longer secure. The U.S. NIST made a step toward mitigating the risk of quantum attacks, by announcing the PQC standardization process [3]. In March 2019, NIST published a list of candidates qualified to the second round of the PQC process [4]. To date, hardware performance of Round 1 candidates was reported for only a small percentage of all submissions.

In this paper, we present the hardware design for low-area implementation of the PQC Round 2 candidate, called Round5. Our design is able to provide both the Key Encapsulation Mechanism (KEM) and Public-Key Encryption (PKE) functionalities. We provide results for all parameter sets of the ring-versions of the respective schemes. Our main goal was to develop the first full-hardware implementation of this PQC submission, able to operate on low-cost FPGAs.

### A. Previous work

Due to complexity of designs, there are only several reports for implementations of PQC candidates in FPGAs. From among lattice-based candidates, Howe et al. [5] reported results for *FrodoKEM*. Kuo et al. [6] and Oder and Güneysu [7] independently reported hardware results for *NewHope*. Aforementioned papers targeted Xilinx Artix-7 FPGA.

In [8], Farahmand et al. proposed a new approach for evaluating PQC candidates by using software/hardware (SW/HW) codesign. They proposed to implement only the most time-consuming functions in the FPGA fabric, while the remaining

parts of the algorithms are implemented in software and run on ARM. Using this SW/HW approach, they reported results for four Round 1 NTRU-based proposals.

For other, non lattice-based candidates, Koziel et al. implemented the isogeny-based SIKE [3], [9]. For multivariate schemes, Ferozपुरi and Gaj reported results for Rainbow [10], implemented using Xilinx Virtex-7 and Kintex-7 FPGAs. From among code-based candidates, Wang et al. reported results for Classic McEliece (a.k.a. classical Niederreiter cryptosystem with binary Goppa codes), implemented using Stratix V FPGAs [11], [12].

### B. Contribution

In this paper, we present a novel hardware design for the ring version of the Round5 submission to the NIST PQC standardization process. Our design is oriented to be a low-area implementation, able to run on low-end FPGAs. The area-performance trade-off is obtained by our customizable architecture for polynomial multiplication.

## II. ROUND5 DESCRIPTION

In the NIST PQC Round 2, there are 26 proposals, with 12 of them belonging to the family of lattice-based schemes. The lattice-based cryptography is a promising option for secure post-quantum KEMs and PKE schemes. It also offers additional novel functionalities, such as homomorphic encryption [13] and identity-based encryption [14].

*Round5* [15] comes from merging two other Round 1 candidates: Round2 [16] and HILA5 [17]. The main underlying problem in *Round5* is Generalized Learning With Rounding (GLWR). In GLWR, the problem randomness comes from rounding, and this feature allows avoiding the necessity of implementing a random bit sampler with any specific distribution, which is a requirement in several other proposals. The submission package contains proposals for indistinguishable under chosen plaintext attack (IND-CPA) KEM and indistinguishable under chosen ciphertext attack (IND-CCA) PKE. Both proposed variants come from the Fujisaki-Okamoto (F-O) transformation [18], by using the main building block of *Round5* — *r5\_cpa\_pke*, the IND-CPA PKE module. Other required modules to perform F-O transformation are a hash function and authenticated encryption with associated data (AEAD). *Round5* comes also in versions with error correcting codes, but our design does not support this functionality.

The package submitted to NIST contains 21 parameters sets, supporting three NIST security levels: 1, 3 and 5. The parameters sets considered in this paper are presented in Table I. We provide results only for the ring version of the schemes, without error correcting codes. The parameter  $n$  describes the polynomial degree,  $p$ ,  $q$ , and  $t$  refer to moduli used in the design for modular reduction and rounding. All moduli are powers of two and must satisfy the requirement  $t < p < q$ .

TABLE I  
PROPOSED PARAMETER SETS FOR THE RING-BASED VERSION OF THE  
ROUND5 PQC CANDIDATE

R5ND Version	$n$	$\log_2(q)$	$\log_2(p)$	$\log_2(t)$
1KEM_0d	618	11	8	4
3KEM_0d	786	13	9	4
5KEM_0d	1018	14	9	4
1PKE_0d	586	13	9	4
3PKE_0d	852	12	9	5
5PKE_0d	1170	13	9	5

#### A. Key generation

In the key generation function, a random seed is expanded by cSHAKE [19]. The use of seed for cSHAKE allows decreasing the size of keys at the expense of an additional cost of expanding the key at the beginning of encryption and decryption. To be compliant with the proposed Hardware API for Post-Quantum Public Key Cryptosystems [20], the key generation function is not a part of the reported design. All long-term keys must be provided to the hardware module from outside, before the main functionality starts.

#### B. Encryption and decryption

Algorithm 1 contains pseudocode for the IND-CPA PKE. The encryption routine starts with expanding a part of the public key and a random input using the cSHAKE function. In the next step, two polynomial multiplications are performed. For polynomial multiplication, one of the polynomials must be lifted to the other's polynomial ring. After the computations, the result is unlifted back. Next, the result is rounded, which is the source of randomness in the GLWR problem.

---

#### Algorithm 1 Round5 Encryption

---

**Require:** public key  $pk$ , message  $msg$ , seed  $\rho$

**Ensure:** ciphertext  $(U, v)$

- 1:  $A \leftarrow \text{Create\_A}(pk.\text{sigma})$
  - 2:  $R \leftarrow \text{Create\_R}(\rho)$
  - 3:  $U \leftarrow \text{Unlift}(\text{Lift}(A) * R)$
  - 4:  $U \leftarrow \text{Round}(U)$
  - 5:  $X \leftarrow \text{Unlift}(\text{Lift}(pk.B) * R)$
  - 6:  $X \leftarrow \text{Round}(X)$
  - 7:  $v \leftarrow msg + X$
  - 8: **return**  $(U, v)$
- 

In Algorithm 2, the pseudocode for decryption is shown. In the first step, the secret key is expanded by using cSHAKE.

Next, only one polynomial multiplication is executed with lifting and unlifting. After the polynomial multiplication, a subtraction from a part of the ciphertext is performed. The last operation is rounding.

---

#### Algorithm 2 Round5 Decryption

---

**Require:** ciphertext  $(U, v)$ , private key  $sk$

**Ensure:** message  $m$

- 1:  $S \leftarrow \text{Create\_S}(sk)$
  - 2:  $X \leftarrow \text{Unlift}(\text{Lift}(U) * S)$
  - 3:  $m \leftarrow v - X$
  - 4:  $m \leftarrow \text{Round}(m)$
  - 5: **return**  $m$
- 

#### C. Supporting functions

Round5 uses three supporting functions during encryption and decryption. First, to obtain an NTRU-like polynomial in the polynomial ring  $\mathbb{Z}_q[x]/(N_{n+1}(x))$  from the key and the random data, the lift function must be applied before multiplication. The lift function is presented in Algorithm 3.

---

#### Algorithm 3 Lift function

---

**Require:** a polynomial  $A$  of length  $n$

**Ensure:** an NTRU-like polynomial  $C$  of length  $n + 1$

- 1:  $C_0 \leftarrow -A_0 \pmod{n}$
  - 2: **for**  $1 \leq i \leq n - 1$  **do**
  - 3:    $C_i \leftarrow A_{i-1} - A_i \pmod{n}$
  - 4: **end for**
  - 5:  $C_n \leftarrow A_{n-1}$ ;
- 

The second function required for proper polynomial multiplication is unlifting, presented in Algorithm 4. Unlifting is applied after polynomial multiplication, and performs polynomial division, taking back the polynomial to  $\mathbb{Z}_q[x]/(\Phi_{n+1}(x))$ .

---

#### Algorithm 4 Unlift function

---

**Require:** a NTRU-like polynomial  $A$  of length  $n + 1$

**Ensure:** a polynomial  $C$  of length  $n$

- 1:  $C_0 \leftarrow -A_0 \pmod{n}$
  - 2: **for**  $1 \leq i \leq n - 1$  **do**
  - 3:    $C_i \leftarrow C_{i-1} - A_i \pmod{n}$
  - 4: **end for**
- 

The last supporting function is rounding, applied to every polynomial coefficient. It is shown in Algorithm 5. It is responsible for rounding elements to smaller values using the exact approach presented in the submission. This function is a source of randomness in the GLWR problem. It is called twice during encryption and once at the end of decryption. Input arguments of rounding are specified in the proposal's documentation.

### III. HARDWARE DESIGN OF ROUND5

We present a low-area architecture of Round5. The implementation follows the proposed hardware API for Post-Quantum Public Key Cryptosystems [20]. The top-level view,

**Algorithm 5** Rounding function

**Require:** an element  $x$  to round, a proper set of  $\{rounding\_constant, shift\_value, mask\}$

**Ensure:** a rounded element  $x$

- 1:  $x \leftarrow x + rounding\_constant$
- 2:  $x \leftarrow x \ll shift\_value$
- 3:  $x \leftarrow x \& mask$
- 4: **return**  $x$

compatible with the aforementioned API, is presented in Fig. 1. The API treats various inputs as public, secret, or random. Thus, three different sets of input ports are used. Each port can handle commands and headers required by the API to control the design.

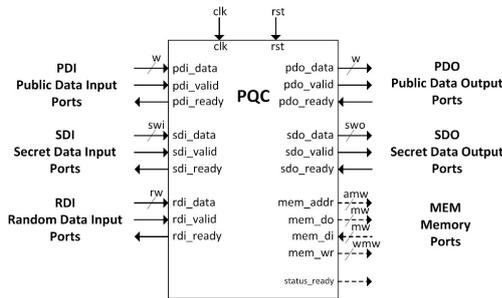


Fig. 1. The PQC Hardware API used in this paper

Going one level down, the architecture of *Round5* is presented in Fig. 2. With the given design, all functionalities of *Round5* are implemented. Each implemented module support all three NIST security levels – 1, 3, and 5. Security level is chosen during compilation and cannot be changed at runtime. The functional modules take input from and write outputs to a shared data bus. The privilege of writing to data bus is granted by the controller’s module.

The main controller is responsible for managing the state of the design and enforcing the proper data flow between modules, depending on a selected operation. It also receives and responds to commands from outside.

The *SHAKE256* module implements the extendable output hash function cSHAKE. It is used for generating pseudo-random polynomials from a given seed. It is also required for the F-O transformation.

The next major component used in our design is the AES-GCM module for authenticated encryption. It is used only in the IND-CCA PKE, as a required part of the F-O transformation. For the IND-CPA KEM, this module can be omitted from the design. The occurrence of AES-GCM is the main difference in the design between the IND-CPA KEM and the IND-CCA PKE versions. In agreement with the F-O transformation, the IND-CCA PKE is build on KEM with additional authenticated encryption.

The most important component is *r5\_cpa\_enc*, providing all arithmetic operations for IND-CPA PKE, the main building

block of all *Round5* proposals.

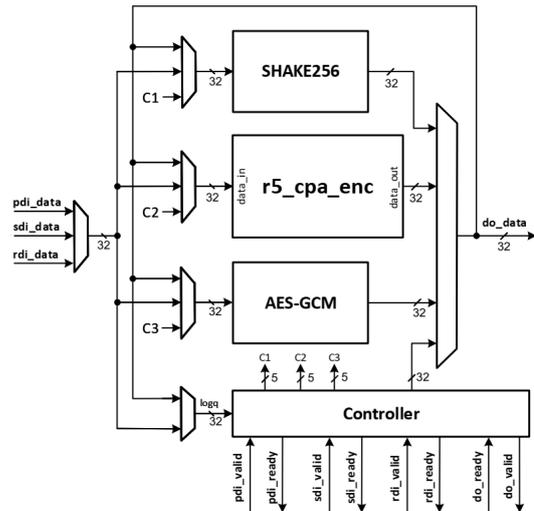


Fig. 2. The proposed top-level architecture of low-area full hardware implementation of Round5

In Fig.3 an arithmetic module responsible for public key encryption and decryption is presented. The important part of this module is the controller, located on the left side of the figure. This controller is responsible for receiving commands from outside, managing the state of the module, and providing proper signals for internal sub-modules. Polynomials required for the operation are stored in separate memory banks. Polynomial multiplication is executed twice during encryption. Thus, two memory banks are able to feed the data to the polynomial multiplier for the first argument. There is some additional memory for the ternary polynomial and for the ciphertext. The ternary polynomial is the same for both multiplications executed during encryption. The message is stored in a register and processed at the end of encryption.

Before polynomial multiplication is executed, one of the polynomials must be lifted to the so called NTRU-ring [21]. This is performed by the *LIFT\_ELEMENT* module, which performs Algorithm 3. Due to the low-area optimization goal, the proposed module lifts elements sequentially, one element at a time. Lifted elements are written back to memory.

The next step is the polynomial multiplication performed by the *POLY\_MUL* module. It requires new data from memory in each cycle. One of the arguments is a coefficient from the lifted polynomial. The second argument is a set of 16 coefficients from the ternary coefficient memory. The module computes results in a sequential fashion, sending further only one coefficient at a time. The first output is ready after  $n$  clock cycles, where  $n$  is a degree of the polynomials. Every next coefficient is ready after  $\lceil n/16 \rceil$  clock cycles.

The computed coefficient follows directly the remaining data path. At first, the coefficient is unlifted to the primary polynomial ring, as shown in Algorithm 4. The next step depends on the operation type. During decryption, subtraction is applied before rounding. During encryption, rounding is

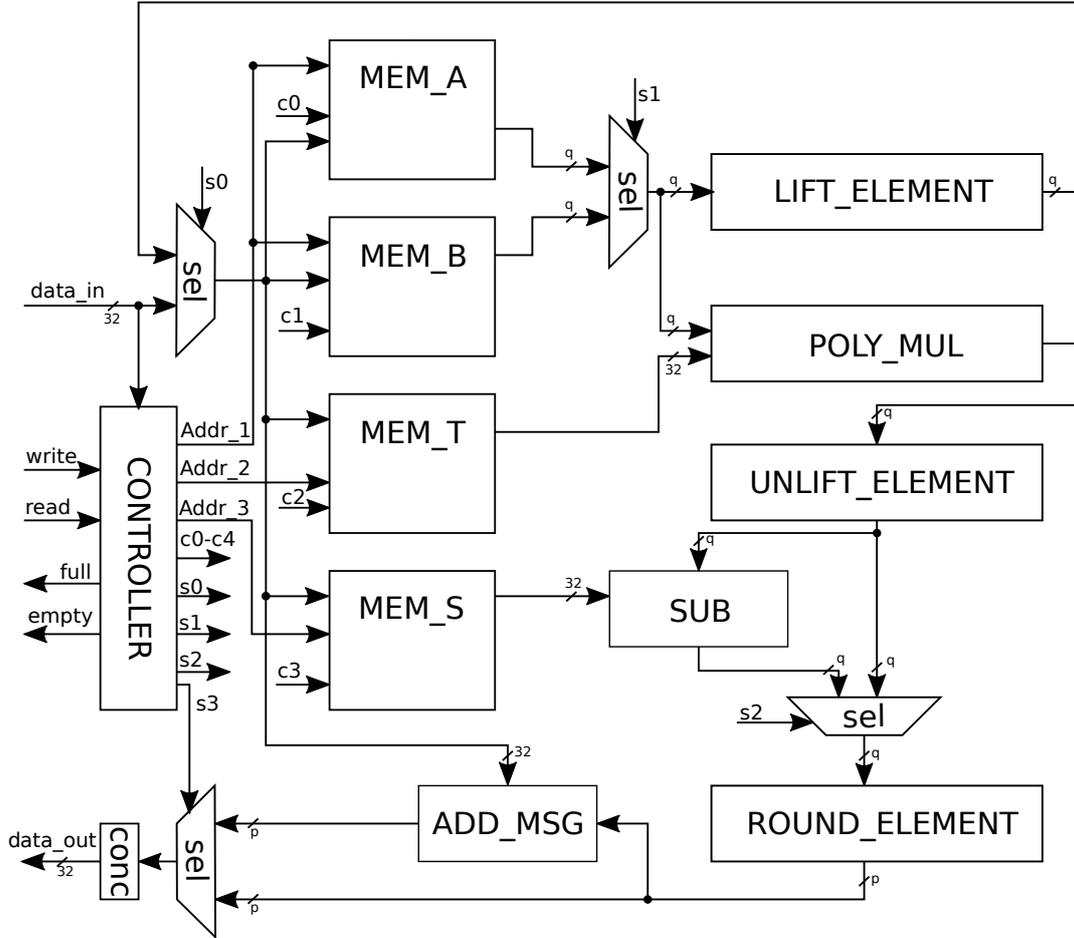


Fig. 3. The arithmetic module for Round5 data encryption and decryption

applied right after unlifting. After rounding, data is stored in the result register directly or with added message bits, depending on the operation type and the state of the encryption process.

Polynomial multiplication has the biggest computational complexity in the *Round5* design. The result is computed by the following formula

$$c_k = \sum_{i+j=k \bmod n} a_i * b_j \quad (1)$$

Due to the form of the polynomial ring and Equation 1, the multiplication can be easily parallelized to speed-up the computations. This is a classical problem of area-performance trade-off, where better performance is achieved by implementing more parallel multipliers, increasing logic usage. Using only one multiplier results in a very large clock cycle latency and slows execution time. On the opposite side of the spectrum, with as many as possible multipliers, the design size is too large to fit in many FPGAs.

We propose a small, in terms of logic utilization, polynomial multiplication module, offering results comparable to those reported for other PQC submissions. Our module executes the

standard schoolbook multiplication with parallel operations. Polynomial multiplication in *Round5* always requires a ternary polynomial and a polynomial with coefficients reduced modulo  $q$ , where  $q$  is a power of 2. Each coefficient in ternary polynomial is from the set  $\{-1, 0, 1\}$ , so only two bits are required to store each coefficient value. All required polynomials are stored separately in internal memory. For polynomials from the ring  $\mathbb{Z}_q[x]$ , each coefficient is accessible directly under different address. Ternary polynomial is stored differently, where one memory cell stores 16 concatenated ternary coefficients. This allows to reduce memory requirements by avoiding only two bits per memory cell utilization. The last memory cell is padded with zeros, if needed.

A new set of coefficients to multiply is loaded from the memory in every clock cycle. The memory address pointers start from the opposite sides, and move in the opposite directions. The ternary polynomial is loaded from the beginning to the end, but the second polynomial is loaded from the last to the first coefficient. The memory pointer for the ternary coefficients is increased by one after each load operation. The second pointer is decreased by the number of parallel multipliers, then the address number is reduced modulo the

polynomial degree. In this scenario, one specific multiplier is computing the same result during one loop over memory. The implemented polynomial multiplication is constant-time.

The proposed polynomial multiplier uses 16 parallel multipliers and is shown in Fig. 4. The number of multipliers is directly linked to the number of coefficients stored in one memory cell. On every memory load, each ternary coefficient is sent to a different multiplier. The second argument is the same for every multiplication unit. The proposed design utilizes the special form of input arguments. The multiplication is done only by addition or subtraction. The value of the accumulator is moved after specific number of multiplications to the next multiplicand. First 15 results are stored in a shift register. This operation is required for computing a proper value. These results are pushed back to multipliers at the end of the computations, to be updated with the remaining multiplication values.

#### IV. RESULTS

We report results for the low-cost DE1-SOC board, manufactured by Terasic. This board is equipped with Intel Cyclone V 5CSEMA5F31C6N FPGA. The chip contains 32,070 adaptive logic modules (ALMs), 128,300 registers, 87 DSP blocks and 3,970 Kb of memory. It contains also the dual-core ARM Cortex-A9 processor. However, in this paper, we focus only on an FPGA part. The post-place and route results were obtained from Intel Quartus Prime v18.1. There is no license requirement for the selected device to perform compilation and deployment.

In Table II, we report results for all security levels and all proposed parameter sets of the IND-CPA KEM and the IND-CCA PKE. We report logic usage for the full design and also for *r5\_cpa\_enc* module separately. This allows us to distinguish the cost of the post-quantum arithmetic module from the remaining costs of the hash-based function and the AEAD module required for the F-O transformation. For presented architecture, the lattice-based arithmetic takes only a small portion of the entire design, several times less than the standard cryptographic elements, such as a hash function or a block cipher. Thus, the implementation goal is achieved. Our design also does not use DSP modules for multiplication, so it can be deployed also with older FPGAs.

The main difference for the logic usage between the IND-CPA KEM and the IND-CCA PKE comes from the additional implementation cost of the AEAD module, not used in the IND-CPA KEM. The logic usage across all security levels is almost the same, as a result of using exactly the same design. The number of multipliers and other arithmetic modules remains always the same.

The memory requirements vary the most among different security levels, due to the necessity of storing significantly larger polynomials. Memory is also used as an input and output buffer to modules in the FIFO queues and in the SHAKE256 implementation. Thus, the overall memory requirements are larger than the sum of all *Round5* elements, such as keys, ciphertext, plaintext, and random data.

All implemented versions of *Round5* run with a similar clock frequency. The reported design is able to perform encapsulation and decapsulation for the highest security level under 1 ms. The encryption and decryption is significantly longer as a result of additional computations required by the F-O transformation and can be performed in around 2 ms also for the highest security level. Only for the lowest security level, the IND-CCA encryption is performed faster than the IND-CPA encapsulation for the parameter set proposed by the submission's authors. These operations are very similar, but for the selected parameter set, the IND-CCA version has smaller polynomial degree. The polynomial degree has the biggest impact on the computational complexity. Thus, faster execution is obtained for the IND-CCA encryption than for the IND-CPA encapsulation.

TABLE II  
OBTAINED RESULTS FOR THE ROUND5 IND-CPA KEM AND THE ROUND5 IND-CCA PKE. \* – RESULTS IN BYTES; \*\* – RESULTS IN CLOCK CYCLES.

<i>Parameter</i>	<i>IND-CPA KEM</i>	<i>IND-CCA PKE</i>	<i>Ratio</i>
<b>Security level: 1</b>			
Parameter set	R5ND_1KEM_0d	R5ND_1PKE_0d	—
PK size*	634	676	1.066
SK size*	16	708	44.25
CT size*	682	754	1.106
Enc latency**	49,714	44,808	0.90
Dec latency**	25,556	67,504	2.64
Total ALMs	4,084	6,305	1.54
Arithm ALMs	448	487	1.08
Memory*	8,445	9,134	1.08
Max freq.	142 MHz	136 MHz	0.95
Enc time	0.35 ms	0.33 ms	0.94
Dec time	0.18 ms	0.50 ms	2.75
<b>Security level: 3</b>			
Parameter set	R5ND_3KEM_0d	R5ND_3PKE_0d	—
PK size*	909	983	1.081
SK size*	24	1,031	42.958
CT size*	981	1,119	1.140
Enc latency**	80,565	94,083	1.17
Dec latency**	41,162	141,546	3.44
Total ALMs	4,098	6,312	1.54
Arithm ALMs	494	467	0.95
Memory*	9,466	10,112	1.06
Max freq.	135 MHz	132 MHz	0.98
Enc time	0.60 ms	0.71 ms	1.19
Dec time	0.30 ms	1.07 ms	3.52
<b>Security level: 5</b>			
Parameter set	R5ND_5KEM_0d	R5ND_5PKE_0d	—
PK size*	1,176	1,349	1.145
SK size*	32	1,413	44.156
CT size*	1,274	1,525	1.197
Enc latency**	132,877	175,965	1.32
Dec latency**	67,556	264,534	3.92
Total ALMs	4,116	6,337	1.54
Arithm ALMs	522	502	0.96
Memory*	10,753	11,765	1.09
Max freq.	133 MHz	130 MHz	0.98
Enc time	1.00 ms	1.35 ms	1.35
Dec time	0.50 ms	2.03 ms	4.01

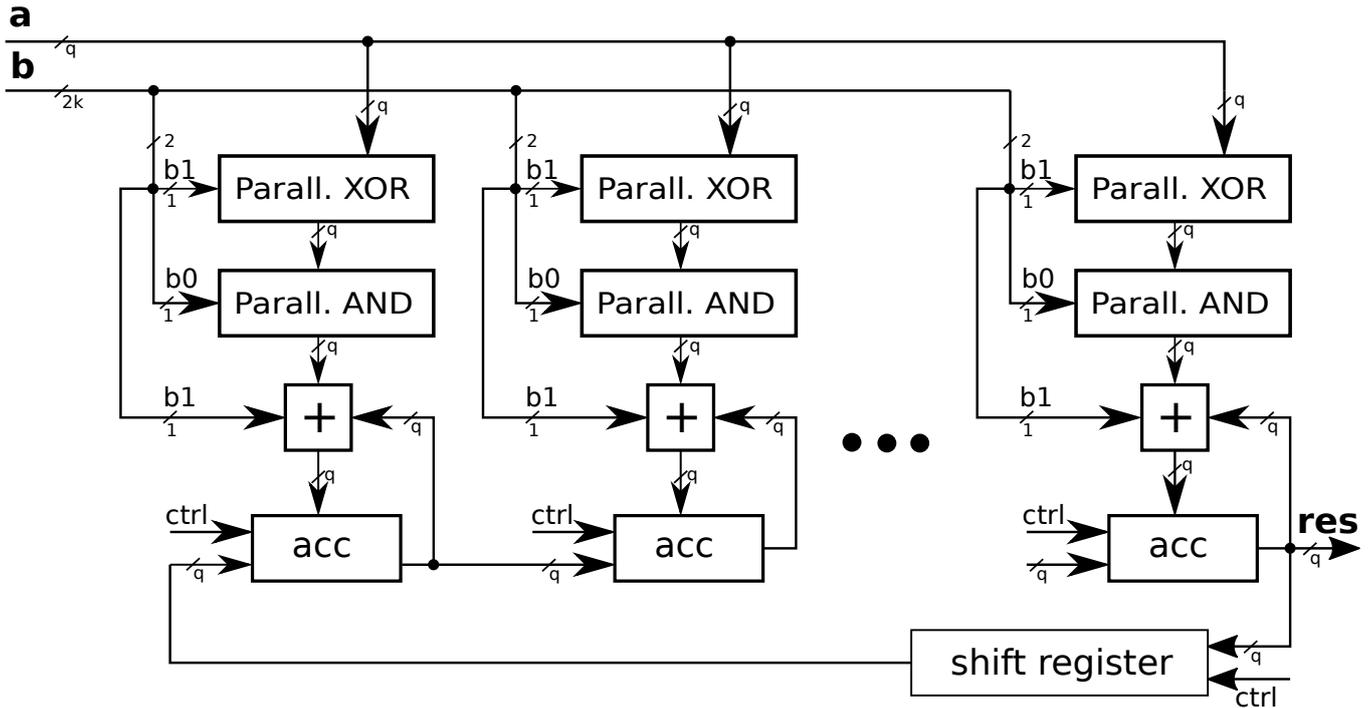


Fig. 4. The parallel polynomial multiplier for Round5

#### A. Comparison to other results

A fair comparison to other results reported to date is hard and complex due to multiple factors directly affecting the obtained results. Moreover, there are no specific guidelines from NIST about proper evaluation of candidates, regarding an FPGA device, implementation goal, API, or compliance criteria. In terms of API and the compliance criteria our design follows the proposal by Ferozpur et al. [20]. As for the FPGA board, we selected one of the least expensive boards, with the free license for the compiler. Any conclusions from comparing logic usage for different FPGA vendors and for different PQC submissions, we are leaving up to the reader.

Howe et al. [5] reported results for full hardware implementation of another lattice-based candidate *FrodoKEM*. They report results for Xilinx Atrix-7 FPGA, and their design balances between area consumption and performance. Their maximum frequency is 167 MHz and is higher than reported in this paper for *Round5*. However, the time required to perform decapsulation is at least an order of magnitude higher, requiring around 20 ms for the execution. Logic requirements are reported for separate modules, not for the entire design able to perform all operations. These modules require around 2,000 slices each.

Most of the other papers reporting results for non-lattice-based PQC candidates, also provide results for Xilinx FPGAs. However, Wang et al. reported results for the Classic McEliece [11], [12] implementation on other high-end Intel FPGA, Stratix V. Their time-optimized implementation uses 121,806 ALM and run at 250 MHz clock, being able to encrypt

and decrypt in less than 0.1 ms.

#### V. CONCLUSIONS AND FUTURE WORK

This paper presented a complete low-area FPGA design for ring version of *Round5*, a lattice-based submission to NIST PQC Standardization process. We reported the post-place and route results for main parameters sets covering all security levels for KEMs and PKEs.

For future work, we consider exploring the area-performance trade-off offered by the proposed polynomial multiplier. A similar polynomial ring is also used by other NTRU-based proposals. Thus, our multiplier can be used for the performance evaluation of other candidates. As for *Round5*, an extension with error correcting codes and the non-ring versions of all schemes is the next big step to provide coverage of all possible parameter sets and versions.

#### ACKNOWLEDGEMENTS

Special thanks to Kris Gaj for his help and valuable comments.

#### REFERENCES

- [1] "Algorithms for quantum computation: Discrete logarithms and factoring."
- [2] R. L. Rivest, A. Shamir, and L. Adleman, "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems," vol. 21, no. 2, pp. 120–126.
- [3] Post-Quantum Cryptography: Call for Proposals. [Online]. Available: <https://csrc.nist.gov/Projects/Post-Quantum-Cryptography/Post-Quantum-Cryptography-Standardization/Call-for-Proposals>
- [4] Post-Quantum Cryptography: Round 2 Submissions. [Online]. Available: <https://csrc.nist.gov/Projects/Post-Quantum-Cryptography/Round-2-Submissions>

- [5] “On Practical Discrete Gaussian Samplers for Lattice-Based Cryptography,” vol. 67.
- [6] P.-C. Kuo, W.-D. Li, Y.-W. Chen, Y.-C. Hsu, B.-Y. Peng, C.-M. Cheng, and B.-Y. Yang, “High Performance Post-Quantum Key Exchange on FPGAs,” p. 17. [Online]. Available: <https://eprint.iacr.org/2017/690.pdf>
- [7] T. Oder and T. Guneyasu, “Implementing the NewHope-Simple Key Exchange on Low-Cost FPGAs,” in *LATINCRYPT 2017*. [Online]. Available: [https://www.ei.ruhr-uni-bochum.de/media/seceng/veroeffentlichungen/2018/04/16/newhope\\_fpga.pdf](https://www.ei.ruhr-uni-bochum.de/media/seceng/veroeffentlichungen/2018/04/16/newhope_fpga.pdf)
- [8] F. Farahmand, V. Dang, D. T. Nguyen, and K. Gaj, “Evaluating the Potential for Hardware Acceleration of Four NTRU-Based Key Encapsulation Mechanisms Using Software/Hardware Codesign.”
- [9] B. Koziel, R. Azarderakhsh, M. Mozaffari Kermani, and D. Jao, “Post-Quantum Cryptography on FPGA Based on Isogenies on Elliptic Curves,” vol. 64, no. 1, pp. 86–99. [Online]. Available: <http://ieeexplore.ieee.org/document/7725935/>
- [10] A. Ferozपुरi and K. Gaj, “High-speed FPGA Implementation of the NIST Round 1 Rainbow Signature Scheme,” in *2018 International Conference on ReConfigurable Computing and FPGAs (ReConFig)*. IEEE, pp. 1–8. [Online]. Available: <https://doi.org/10.1109/reconfig.2018.8641734>
- [11] W. Wang, J. Szefer, and R. Niederhagen, “FPGA-based Key Generator for the Niederreiter Cryptosystem Using Binary Goppa Codes,” in *Cryptographic Hardware and Embedded Systems – CHES 2017*, W. Fischer and N. Homma, Eds. Springer International Publishing, vol. 10529, pp. 253–274. [Online]. Available: [https://doi.org/10.1007/978-3-319-66787-4\\_13](https://doi.org/10.1007/978-3-319-66787-4_13)
- [12] —, “FPGA-Based Niederreiter Cryptosystem Using Binary Goppa Codes,” in *PQCrypto 2018*, ser. LNCS, T. Lange and R. Steinwandt, Eds., vol. 10786. Springer International Publishing, pp. 77–98. [Online]. Available: [https://doi.org/10.1007/978-3-319-79063-3\\_4](https://doi.org/10.1007/978-3-319-79063-3_4)
- [13] C. Gentry, “Fully homomorphic encryption using ideal lattices,” in *Proceedings of the 41st Annual ACM Symposium on Symposium on Theory of Computing - STOC '09*. ACM Press, p. 169. [Online]. Available: <https://doi.org/10.1145/1536414.1536440>
- [14] T. Guneyasu and T. Oder, “Towards lightweight Identity-Based Encryption for the post-quantum-secure Internet of Things,” in *2017 18th International Symposium on Quality Electronic Design (ISQED)*. IEEE, pp. 319–324. [Online]. Available: <https://doi.org/10.1109/ISQED.2017.7918335>
- [15] I. T. L. Round5 Submission Team. Round 2 Submissions -Round5 candidate submission package. [Online]. Available: <https://csrc.nist.gov/CSRC/media/Projects/Post-Quantum-Cryptography/documents/round-2/submissions/Round5-Round2.zip>
- [16] H. Baan, S. Bhattacharya, O. Garcia-Morchon, R. Rietman, L. Tolhuizen, J.-L. Torre-Arce, and Z. Zhang, “Round2: KEM and PKE based on GLWR,” p. 72.
- [17] M.-J. O. Saarinen, “HILA5: On Reliability, Reconciliation, and Error Correction for Ring-LWE Encryption,” pp. 192–212. [Online]. Available: [https://doi.org/10.1007/978-3-319-72565-9\\_10](https://doi.org/10.1007/978-3-319-72565-9_10)
- [18] E. Fujisaki and T. Okamoto, “Secure Integration of Asymmetric and Symmetric Encryption Schemes,” vol. 26, no. 1, pp. 80–101. [Online]. Available: <https://doi.org/10.1007/s00145-011-9114-1>
- [19] J. Kelsey, S.-j. Chang, and R. Perlner, “SHA-3 derived functions: cSHAKE, KMAC, TupleHash and ParallelHash.” [Online]. Available: <https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-185.pdf>
- [20] A. Ferozपुरi, F. Farahmand, V. B. Dang, M. U. Sharif, J.-P. Kaps, and K. Gaj, “Hardware API for Post-Quantum Public Key Cryptosystems.” [Online]. Available: [https://cryptography.gmu.edu/athena/PQC/PQC\\_HW\\_API.pdf](https://cryptography.gmu.edu/athena/PQC/PQC_HW_API.pdf)
- [21] J. Hoffstein, J. Pipher, and J. H. Silverman, “NTRU: A ring-based public key cryptosystem,” in *Algorithmic Number Theory*, J. P. Buhler, Ed. Springer Berlin Heidelberg, vol. 1423, pp. 267–288. [Online]. Available: <https://doi.org/10.1007/BFb0054868>



# Accelerating Multivariate Cryptography with Constructive Affine Stream Transformations

Michael Carenzo  
Rochester Institute of Technology  
1 Lomb Memorial Dr,  
Rochester, NY 14623  
Email: mec2487@rit.edu

Monika Polak  
Rochester Institute of Technology  
1 Lomb Memorial Dr,  
Rochester, NY 14623  
Email: mkpvcs@rit.edu

**Abstract**—On December 20th, 2016, the National Institute of Standards and Technology (NIST) formally initiated a competition to solicit, evaluate, and standardize one or more quantum-resistant cryptographic algorithms. Among the current candidates is a cryptographic primitive which has shown much promise in the post-quantum age, Multivariate Cryptography. These schemes compose two affine bijections  $S$  and  $T$  with a system of multivariate polynomials. However, this composition of  $S$  and  $T$  becomes costly as the data encrypted grows in size. Here we present Constructive Affine Stream (CAS) Transformations, a set of algorithms which enable specialized, large-scale, affine transformations in  $O(n)$  space and  $O(n \log n)$  time, without compromising security. The goal of this paper is to address the practical problems related to affine transformations common among almost all multivariate cryptographic schemes.

## I. INTRODUCTION

MULTIVARIATE Cryptography is a, post-quantum, cryptographic primitive based on the difficulty of solving systems of multivariate equations over a finite field [1]. At their core, multivariate schemes define a set of (usually quadratic) polynomials:

$$\begin{pmatrix} p_1(w_1, \dots, w_n) \\ \dots \\ p_m(w_1, \dots, w_n) \end{pmatrix}$$

where all coefficients and variables are in  $\mathbb{F}_q$ , a field with  $q$  elements. Given a plaintext message:  $(x_1, \dots, x_n) \in \mathbb{F}_q^n$  the ciphertext is computed by evaluating:

$$\mathcal{P}(x_1, \dots, x_n) = \begin{pmatrix} p_1(x_1, \dots, x_n) \\ \dots \\ p_m(x_1, \dots, x_n) \end{pmatrix} = \begin{pmatrix} c_1 \\ \dots \\ c_m \end{pmatrix}$$

To decrypt the ciphertext  $(c_1, \dots, c_m)$ , one must hold the *secret key* used to generate the polynomials in  $\mathcal{P}$  in order to invert  $\mathcal{P}$ .

$$\mathcal{P}^{-1}(c_1, \dots, c_m) = (x_1, \dots, x_n)$$

Inverting  $\mathcal{P}$  without the secret key is equivalent to solving a system of multivariate equations over a finite field, known formally as the  $\mathcal{MQ}$ -Problem, and is proven to be NP-Hard. However, modern constructions of these schemes rarely use a set of multivariate polynomials on their own. Modern constructions almost always compose the set of polynomials with two invertible affine maps  $S$  and  $T$  [2]. So, in reality:

$$\mathcal{P} = T \circ Q \circ S : \mathbb{F}_q^n \rightarrow \mathbb{F}_q^m$$

Where  $Q$  (also known as the *central* or *core map*) is the set of multivariate polynomials and  $S$  and  $T$  are (sometimes linear) affine maps of full-rank. While many papers focus on the design of the central map, few describe how to effectively generate and compose  $S$  and  $T$ , despite the importance of this operation [3], [4], [5], [6], [7]. As the plaintext grows in size, so too do  $S$  and  $T$ . In fact,  $S$  and  $T$  grow so fast that their composition becomes intractable very quickly.

This poses a significant hurdle for symmetric applications of multivariate cryptography. Consider encrypting a 1kB, 500kB, and 1MB file. Because  $S$  and  $T$  are  $n \times n$  matrices where  $n$  is the size of the plaintext, these files require matrices 1MB ( $1\text{kB}^2$ ), 250GB ( $500\text{kB}^2$ ), and 1TB ( $1\text{MB}^2$ ) in size. This rapid inflation of  $S$  and  $T$  restricts these schemes from (reasonably) encrypting anything larger than  $\sim 1\text{kB}$  (without chaining).

In this paper we present Constructive Affine Stream (CAS) Transformations, a set of algorithms capable of efficiently generating and multiplying  $S$  and  $T$  by any arbitrary vector. At the same time, these transformations preserve the post-quantum security of multivariate ciphers. We begin by presenting the theory behind these transformations (Sec. II), followed by a general implementation (Sec. III). We then analyze the asymptotics of the aforementioned implementation (Sec. IV) and conclude by evaluating how these transformations impact the security of multivariate ciphers (Sec. V).

## II. CONSTRUCTIVE AFFINE STREAM TRANSFORMATIONS

Instead of generating, storing, and operating on a matrix outright, a Constructive Affine Stream (CAS) deterministically generates (via some seed derived from the private key) a stream of integers that represent an affine matrix of full-rank. These streams can then be used to transform a vector progressively, in the same way a normal matrix-vector multiplication would, without ever having to store an actual matrix. Furthermore, a stream (if configured with the same seed) can be switched to generate the inverse of a given matrix so that a previous transformation can be undone.

### A. The Structure of an Affine Stream

Constructive Affine Streams (CAS) leverage a basic property of matrix-vector multiplication; each (matrix) row is dotted with the vector term, one at a time, independently of

the other rows. In effect, this property allows the values of each row to be randomly generated over the course of its dot product with the vector term and then “thrown away.” This randomly generated sequence of values is what an affine stream is composed of and allows a vector to be transformed without having to store a matrix.

However, randomly generating matrix rows doesn’t guarantee that the resulting matrix is invertible. Furthermore, even if the matrix was invertible, the values of the rows are generated as needed and never stored, thus each matrix value is “blind” to the values adjacent to it. This limitation prevents a matrix’s inverse from being computed using conventional techniques. In order to solve this problem, matrices generated by a CAS maintain a specific structure.

A CAS generated matrix:

- 1) Takes the form of an upper or lower triangular matrix;
- 2) With non-zero values on the main-diagonal; and
- 3) For every row/column pair which intersect on the main-diagonal, only one of the two (row or column) can contain non-zero values.

While conditions 1 and 2 ensure that the matrix stream produced is always invertible, condition 3 guarantees that the matrix stream is invertible, one row at a time, using only values on the main-diagonal. While the result of condition 3 may not seem intuitive at first, consider inverting the matrix in Fig. 1 via Gauss-Jordan elimination.

A row that only contains zeros (excluding the main-diagonal) can only eliminate values down the column which intersects it on the main-diagonal. For instance, in Fig. 1,  $r_1$  will only be used to eliminate values in  $c_1$ . A row that contains multiple non-zero values has nothing to eliminate in the column which intersects it on the main-diagonal because, by the definition above, that column will always contain zeros. Returning to Fig. 1,  $r_3$  will never need to eliminate anything down  $c_3$  because it is guaranteed to be “zero-valued” by definition. However,  $r_3$ ’s non-zero values 5 and 4 will be eliminated by rows  $r_0$  and  $r_1$  respectively.

Described more generally, the only rows which perform elimination are the ones which contain a single non-zero value: the diagonal-component. Moreover, the only columns which contain non-zero values are the columns which intersect one of the aforementioned rows on its diagonal-component. Consequently, each column can be eliminated independently, without altering adjacent columns.

Because every column can be eliminated on its own, it follows that every column can be inverted on its own. This result is what enables on-the-fly, row-by-row, CAS inversion. The only information needed beforehand are the values of the main-diagonal. With these values, each row can be inverted, one at a time, by mapping its diagonal-component to its multiplicative inverse and all other values (off the main-diagonal) via:

$$-\alpha_{i,j} \times \alpha_{j,j}^{-1} \times \alpha_{i,i}^{-1}$$

where  $\alpha_{i,j}$  is the matrix value at the  $i$ th column and  $j$ th row.

Form	5×5 CAS
Matrix Representation	$\begin{matrix} & c_0 & c_1 & c_2 & c_3 & c_4 \\ r_0 & \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \end{bmatrix} \\ r_1 & \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \end{bmatrix} \\ r_2 & \begin{bmatrix} 8 & 0 & 1 & 0 & 0 \end{bmatrix} \\ r_3 & \begin{bmatrix} 5 & 4 & 0 & 1 & 0 \end{bmatrix} \\ r_4 & \begin{bmatrix} 0 & 7 & 0 & 0 & 1 \end{bmatrix} \end{matrix}$
Stream	[1,0,8,5,0,1,0,4,7,1,1,1]

Fig. 1. An Example CAS Stream

The derivation of this equation is fairly straight forward. Given the following augmented matrix:

$$\left[ \begin{array}{ccc|ccc} \alpha_{i,i} & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \alpha_{i,j} & \cdots & \alpha_{j,j} & 0 & \cdots & 1 \end{array} \right]$$

we can solve for the inverse matrix value at position  $\alpha_{i,j}$ , via Gauss-Jordan elimination. Here, the rows at  $i$  and  $j$  are normalized:

$$\left[ \begin{array}{ccc|ccc} 1 & \cdots & 0 & \alpha_{i,i}^{-1} & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \alpha_{i,j} \times \alpha_{j,j}^{-1} & \cdots & 1 & 0 & \cdots & \alpha_{j,j}^{-1} \end{array} \right]$$

then the row at  $i$  (multiplied by  $\alpha_{i,j} \times \alpha_{j,j}^{-1}$ ) is subtracted from the row at  $j$ :

$$\left[ \begin{array}{ccc|ccc} 1 & \cdots & 0 & \alpha_{i,i}^{-1} & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & -\alpha_{i,j} \times \alpha_{j,j}^{-1} \times \alpha_{i,i}^{-1} & \cdots & \alpha_{j,j}^{-1} \end{array} \right]$$

Note that in Fig. 1, the main-diagonal only contains ones. This matrix configuration is known formally as a “Semi-Byte” CAS and is one of the three main stream types. A Semi-Byte CAS is trivial to invert because the main-diagonal doesn’t need to be generated beforehand (as it is already known) and each value off of the main-diagonal is mapped via:

$$-\alpha_{i,j} \times 1^{-1} \times 1^{-1} = -\alpha_{i,j}$$

### B. Performing Constructive Affine Stream Transformations

Notice that the affine stream in Fig. 1 is not a consecutive series of matrix rows. These streams take advantage of the aforementioned matrix constraints to generate only the values necessary for vector transformation. Structurally, affine streams are organized by column and do not contain the zeroed half of the triangular matrix. Furthermore, they do not contain the zeros of zero-valued columns, only their diagonal component. For example, in Fig. 1, 8 was generated in  $r_2$ , therefore  $c_2$  must be a zero-valued column. The only value in the stream from  $c_2$  is its diagonal component: 1.

With these streams, a CAS Transformation can be applied to a vector (*vector*) by iterating over a CAS stream and, for each stream value  $v$ , we apply:

$$\text{transform}[r_v] \leftarrow \text{transform}[r_v] + (\text{vector}[c_v] \times v)$$

Where *transform* is an empty vector which stores the transformation and  $r_v$  and  $c_v$  are the row/column matrix coordinates of  $v$ . (e.g. In Fig. 1, 8 has the  $(r_v, c_v)$  coordinates  $(2, 0)$ .)

### III. IMPLEMENTATION

There are three (main) types of CAS: Binary, Semi-Byte and Byte. Each type requires methods for deterministically generating random bits and random non-zero numbers. The bits determine where non-zero values are placed in a matrix while the numbers determine what the values are. If the sequence of bits and numbers can't be regenerated, a transformation can't be inverted. For the sake of our experiments, we leveraged Trivium ([8]) for our bit generator and a simple linear congruential generator (LCG) for our number generator.

Note that in the implementation that follows: *rand* refers to some Pseudo-Random Number Generator (PRNG) initialized with a password which seeds the bit and number generator. (This password can be the same one used for multivariate encryption.) *lowerTriangular* is a Boolean indicating whether the generated CAS matrix is upper or lower triangular. Lastly, the operators  $*$  and  $+$  refer to group multiplication and addition specific to the chosen finite field.

#### A. Semi-Byte CAS Transformations

A Semi-Byte CAS generates an affine stream composed of 1's down the main-diagonal and 8-bit values everywhere else. Despite its name, matrix values aren't limited to 8-bits and should operate in whatever finite field is selected for encryption. (e.g.  $GF(2^8) \rightarrow$  8-bit matrix values,  $GF(2^{16}) \rightarrow$  16-bit matrix values) To save space, the implementation described below performs the vector transformation in-place. However, it could easily be modified to store the values in a new vector.

---

#### Algorithm 1 Semi-Byte CAS Transformation of a *vect*

---

```

1: emptyColumns  $\leftarrow$  [0] * vect.length
2: for  $i = 0$  to vect.length do
3:   if emptyColumns[ $i$ ] == 0 then
4:     for  $j = 1$  to vect.length -  $i$  do
5:       if rand.getBit() == 1 then
6:         scalar  $\leftarrow$  rand.getBytes()
7:         if lowerTriangular then
8:           vect[ $i + j$ ]  $\leftarrow$  vect[ $i + j$ ] + (vect[ $i$ ] * scalar)
9:         else
10:          vect[ $i$ ]  $\leftarrow$  vect[ $i$ ] + (vect[ $i + j$ ] * scalar)
11:        end if
12:        emptyColumns[ $i + j$ ]  $\leftarrow$  1
13:      end if
14:    end for
15:  end if
16: end for

```

---

In order to perform an inverse transformation (relative to a given seed) this implementation would be altered as follows:

- Line 8:  $\text{vect}[i + j] \leftarrow \text{vect}[i + j] - (\text{vect}[i] * \text{scalar})$
- Line 10:  $\text{vect}[i] \leftarrow \text{vect}[i] - (\text{vect}[i + j] * \text{scalar})$

#### B. Binary CAS Transformations

A Binary CAS generates an affine stream composed of only 0's and 1's. Its implementation is practically identical to that of a Semi-Byte CAS Transformation. However, in the case of Binary CAS Transformations, *scalar* is always equal to 1.

#### C. Byte CAS Transformations

A Byte CAS generates an affine stream composed of 8-bit values. (Again, bear in mind that values aren't necessarily fixed to 8-bits and, in reality, are bound by the chosen finite field.) These transformations cannot be done in-place and are the costliest in-terms of space-complexity. Implementation details regarding this transformation type can be found at [9].

### IV. ASYMPTOTIC ANALYSIS & PERFORMANCE

Compared to typical matrix-vector multiplication, CAS Transforms are quite efficient. In terms of space complexity, both Binary and Semi-byte CAS Transforms require only  $\Theta(n)$  space to store the columns flagged as zero-valued (*emptyColumns*). Byte CAS Transforms require  $2n$  space for transformations and  $3n$  for inversions, giving both operations a lower-bound of  $\Omega(n)$ .

Computing time-complexity is slightly more complicated due to the probability involved in CAS generation. While normal matrix-vector multiplication is an  $n^2$  operation (for square matrices), CAS Transformations effectively "skip" zero-valued columns as they only operate on their diagonal component. Because these columns are skipped, we can compute the average upper-bound of a CAS Transform by multiplying the height of the columns by the average number of non-zero columns. However, this requires a function which can approximate the average number of non-zero columns a CAS will generate given a matrix size.

While there are several approaches that could be used to derive this function, we chose a statistical approach as it seemed to yield the best estimates. This approach involved randomly generating 5000 (Binary) CAS matrices at each size ranging from  $1 \times 1$  to  $3000 \times 3000$  and computing the average number of non-zero columns at each size. Each matrix was generated using its own instantiation of Trivium initialized with the first 160-bits of a SHA-256 hash derived from a (randomly generated) alphanumeric password. Plotting these results, it appears that the number of non-zero columns grows in a logarithmic fashion. Indeed, once we fitted a logarithmic curve to the data we found that it fit perfectly (Fig. 2a).

This would seem to indicate that the runtime of a CAS Transformation is on average  $O(n \log n)$ . To verify this result, we plotted the average length of a CAS stream. Using the same parameters, (5000 trials over 3000 sizes) we found that the average length of a CAS stream perfectly fits an  $n \log n$  curve (Fig. 2b). Note, this run-time applies to all CAS types because all types generate their matrix structure using the same bit generation algorithm (in our case, via Trivium).

With the proper implementation, these transformations have the potential to reach impressive speeds and are limited primarily by their chosen bit and number generation algorithms.

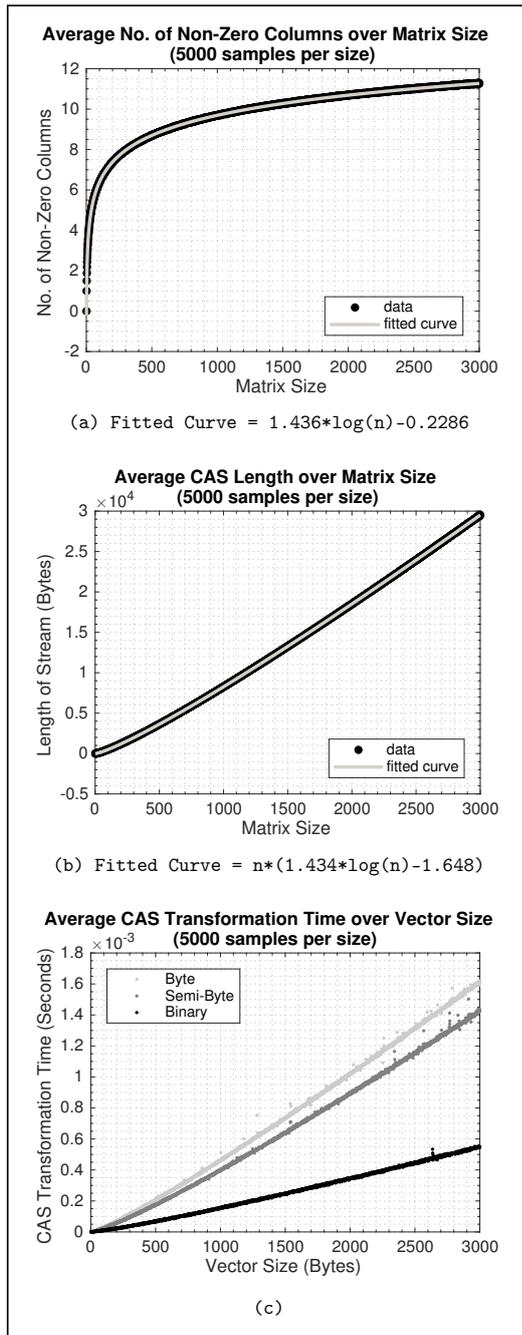


Fig. 2. Runtime Statistics &amp; Time Plots

Our naïve C implementation (run on a computer with an Intel Core i5-3570 processor running at 3.40GHz using 16GB of RAM, with Windows 10 Pro) achieved respectable speeds on its own (Fig. 2c). However, it could be made to operate even faster with an optimized Trivium implementation. In theory a CAS stream could even be cached and applied over blocks. Observe that our speed tests further support our time-complexity analysis as the speed plots reflect our time-complexity plot.

## V. SECURITY

In practice, one CAS Transformation alone isn't enough to sufficiently "mix-up" an input vector. In fact, any triangular matrix on its own isn't enough. The use of single triangular matrices can even break certain multivariate schemes. For example, the multivariate scheme based on the family of expander graphs  $D(n, q)$  is rendered totally insecure when  $S$  and  $T$  take the form of lower-triangular matrices. (For more details about this family of graphs see [10].) To illustrate this insecurity, consider Example 5.1.

*Example 5.1 (The "Poor Mixing" Vulnerability):*

$$\text{Plaintext} = [x_1, x_2, x_3, x_4]$$

$$\text{Password} = [2, 1, 10, 5]$$

$$T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix}, \quad S = T^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

(Note: Often  $S = T^{-1}$  but it is not a general rule.)

The resulting polynomials produced using  $S$ ,  $T$ , and the polynomials generated from  $D(n = 4, q)$  graphs (traversed with *Password*) are:

$$p_1(x_1) = x_1 + 18$$

$$p_2(x_1, x_2) = -18x_1 + x_2 - 201$$

$$p_3(x_1, x_2, x_3) = -18x_1^2 - 201x_1 - 18x_2 + x_3 - 513$$

$$p_4(x_1, x_2, x_3, x_4) = 36x_1^2 + 621x_1 + 18x_2 + x_4 + 3261$$

For the sake of simplicity, these polynomials are not bound to any specific finite field  $GF(q)$ . (All work is shown at [9].)

Notice that each subsequent polynomial introduces one new variable ( $x_i$ ). So, given the piece of ciphertext  $c_1$  produced by  $p_1(x_1)$ , it would be trivial to compute the plaintext piece  $x_1$  by solving  $x_1 = c_1 - 18$ . Then, given  $x_1$ , we could solve for  $x_2$  in  $p_2(x_1, x_2)$  by plugging in  $x_1$  and  $c_2$ . This process can be repeated until all  $x_i$ 's have been solved for and the plaintext message is revealed (without use of the private key).

To eliminate this vulnerability, a lower and upper triangular CAS Transformation can be combined to form a "Square CAS Transformation." This is simply achieved by applying an upper and lower triangular CAS Transformation to a vector, in any order. Because matrix multiplication is associative, this is equivalent to multiplying a vector by the product of an upper and lower triangular CAS matrix. In fact, simulations have shown that a combined upper and lower CAS Transformation correspond to multiplication by a matrix  $A$ , such that for all  $i \neq j$   $P(a_{i,j} = 0) \approx 0.65$  and  $i = j$   $P(a_{i,j} = 0) \approx 0.35$  [9].

Illustrated below are the corrected polynomials which leverage an upper and lower triangular matrix for both  $S$  and  $T$ :

$$p_1(x_1, x_2, x_3, x_4) = -54x_1^2 - 216x_1x_2 - 108x_1x_3 - \dots$$

$$p_2(x_1, x_2, x_3, x_4) = 18x_1^2 + 72x_1x_2 + 36x_1x_3 + \dots$$

$$p_3(x_1, x_2, x_3, x_4) = -18x_1^2 - 72x_1x_2 - 36x_1x_3 - \dots$$

$$p_4(x_1, x_2, x_3, x_4) = 36x_1^2 + 144x_1x_2 + 72x_1x_3 + \dots$$

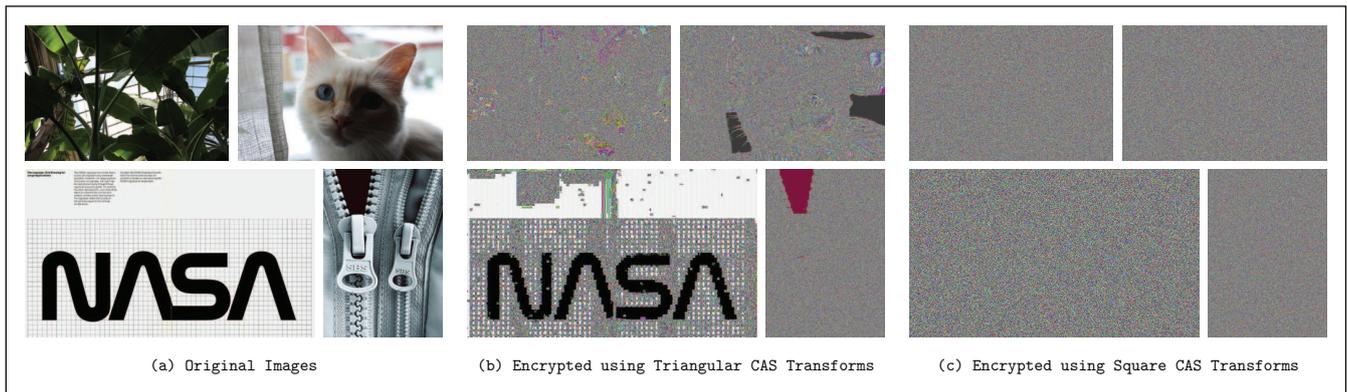


Fig. 3. Image Encryption Tests

A Square CAS Transform can easily be inverted by applying the inverse of its lower and upper triangular components in reverse order. That is, given a vector  $\sigma$ , its transformation  $\sigma'$ , and a triangular CAS Transform  $T(x)$ :

$$T_{lower}(T_{upper}(\sigma)) = \sigma', \quad T_{upper}^{-1}(T_{lower}^{-1}(\sigma')) = \sigma$$

We can visualize this “poor mixing” vulnerability by encrypting image data. In Fig. 3, the images in (a) are composed with an upper-triangular (Semi-Byte) CAS  $T$ , passed through a set of polynomials (from  $D(n, q)$ ), and then composed with another upper-triangular CAS  $S$  to produce the images in (b). Clearly, a large amount of information is exposed. However, if we utilize a square (Semi-Byte) CAS for both  $S$  and  $T$  we produce the images in (c). (Additional tests with alternate configurations can be found at [9].)

While in theory more than two CAS Transformations could be applied to a single vector, Square CAS Transforms on their own have shown to be sufficient.

To further validate the security of Square CAS Transforms, we analyzed the average order of each (square) CAS type. This statistic is of particular interest as there is theoretical evidence which suggests that large order matrices provide better security [11]. (The order of a matrix  $M_{n \times n}$  over a finite field is the smallest positive integer  $k$  such that  $M^k = I_n$ .) At present, the resources required to compute  $k$  over large matrices has restricted our ability to collect large samples of data. However, our initial tests indicate that the order of a Binary matrix rarely exceeds  $k = 120$ . In contrast, Semi-Byte and Byte matrices have both exhibited  $k$ 's in the millions and even billions [9]. This suggests what one might already expect; Semi-Byte and Byte Transforms are considerably stronger than Binary Transforms, making them the preferred choice for most implementations.

## VI. CONCLUSION

As quantum computers march closer towards their full realization we must be prepared with a new set of cryptographic primitives which remain secure in the post-quantum age. Multivariate Cryptography is one of the handful of primitives which has shown real promise but still requires more research

and development before it can be realistically implemented. In this paper we formalized the notion of CAS Transformations, a set of algorithms capable of performing specialized, large-scale, affine transformations in  $O(n)$  space-complexity and  $O(n \log n)$  time-complexity. While there is still more to be learned about these transformations, they have demonstrated real potential; capable of increasing the speed and scale at which Multivariate Cryptography can be applied.

## REFERENCES

- [1] D. J. Bernstein, *Introduction to post-quantum cryptography*. Springer, Berlin, Heidelberg, 2009. [Online]. Available: doi.org/10.1007/978-3-540-88702-7\_1
- [2] J. Ding and B. Y. Yang, *Multivariate Public Key Cryptography*. Springer, Berlin, Heidelberg, 2009. [Online]. Available: doi.org/10.1007/978-3-540-88702-7\_6
- [3] J. Ding, M. S. Chen, A. Petzoldt, D. Schmidt, and B. Y. Yang. (2019) Rainbow digital signature algorithm, nist post-quantum cryptography project, round 2 submissions. [Online]. Available: https://csrc.nist.gov/Projects/Post-Quantum-Cryptography/Round-2-Submissions
- [4] K. Sakumoto, T. Shirai, and H. Hiwatari, *Public-Key Identification Schemes Based on Multivariate Quadratic Polynomials*. Springer, Berlin, Heidelberg, 2011. [Online]. Available: doi.org/10.1007/978-3-642-22792-9\_40
- [5] C. Tao, H. Xiang, A. Petzoldt, and J. Ding, “Simple matrix - a multivariate public key cryptosystem (mpkc) for encryption,” *Finite Fields and Their Applications*, 2015. doi: 10.1016/j.ffa.2015.06.001. [Online]. Available: doi.org/10.1016/j.ffa.2015.06.001
- [6] M. Polak, U. Romańczuk, V. Ustimenko, and A. Wróblewska, “On the applications of extremal graph theory to coding theory and cryptography,” *Electronic Notes in Discrete Mathematics*, 2013. doi: 10.1016/j.endm.2013.07.05. [Online]. Available: doi.org/10.1016/j.endm.2013.07.051
- [7] V. Ustimenko, U. Romańczuk-Polubiec, A. Wróblewska, M. Polak, and E. Zhupa, “On the constructions of new symmetric ciphers based on nonbijective multivariate maps of prescribed degree,” *Security and Communication Networks*, 2019. doi: 10.1155/2019/2137561. [Online]. Available: doi.org/10.1155/2019/2137561
- [8] C. D. Cannière and B. Preneel, *Trivium*. Springer, Berlin, Heidelberg, 2008. [Online]. Available: doi.org/10.1007/978-3-540-68351-3\_18
- [9] M. Careno. (2019) Study of multivariate cryptography, rit independent study website. [Online]. Available: https://www.cs.rit.edu/~mcc2487
- [10] F. Lazebnik, V. A. Ustimenko, and A. J. Woldar, “A new series of dense graphs of high girth,” *Bull. Amer. Math. Soc.*, 1995. doi: 10.1090/S0273-0979-1995-00569-0. [Online]. Available: doi.org/10.1090/S0273-0979-1995-00569-0
- [11] V. Ustimenko, “On the families of stable multivariate transformations of large order and their cryptographical applications,” *Tatra Mountains Mathematical Publications*, 2018. doi: 10.1515/tmmp-2017-0021. [Online]. Available: doi.org/10.1515/tmmp-2017-0021



# Spline-Wavelet Bent Robust Codes

Alla Levina,  
University ITMO, ul. Lomonosova 9  
and  
St. Petersburg Electrotechnical  
University "LETI", Professora Popova  
str., 5  
Russia, Saint-Petersburg,  
Email: alla\_levina@mail.ru

Gleb Ryaskin,  
University ITMO, Saint-Petersburg,  
Russia.  
Email: ryaskingleb20@gmail.com

Igor Zikratov  
The Bonch-Bruевич Saint-Petersburg  
State University of  
Telecommunications, Saint-  
Petersburg, Russia.  
Email: igzikratov@yandex.ru

**Abstract.** This paper presents an application of spline-wavelet transformation and bent-functions for the construction of robust codes. To improve the non-linear properties of presented robust codes, bent-functions were used. Bent-functions ensure maximum non-linearity of functions, increasing the probability of detecting an error in the data channel. In the work different designs of codes based on wavelet transform and bent-functions are developed. The difference of constructions consists of using different grids for wavelet transformation and using different bent-functions. The developed robust codes have higher characteristics compared to existing. These codes can be used for ensuring the security of transmitted information.

## I. INTRODUCTION

Wavelet transformation has become well known and widely used in many fields of science [1-3]. The basic concepts of wavelet theory can be found in the work of Daubechies [3]. Also, wavelet theory has found implementation in the technical fields such as data compression, signal analysis, and communication applications [4-7].

One new direction of implementing wavelet theory is error protection codes [6-7, 9-15]. Error detecting codes are used for the protection in telecommunication channels, they ensure the reliability and security of devices from soft, hard errors and side channel attacks [17]. The purpose of error correcting codes is to provide digital communication over the channel in such a way that errors in the transmission of bits can be detected and corrected by the receiver. This goal is achieved by using coding algorithms that convert information before sending it [16-19].

By exercising various effects on the hardware component of a cryptographic device in order to cause distortion of information at some stages of coding, managing and analyzing errors, an attacker can change the information transmitted over the channel. This type of attack is called a calculation error attack [17]. To provide protection against this type of attack, robust codes built on non-linear functions are used because linear functions do not show all errors due to linear properties [16-17]. And often the most interesting are non-linear functions for which the property of non-linearity is bent function [8].

In this article, was investigated the properties of robust codes constructed on bent functions and wavelet decompositions. Will be presented various methods for constructing this class of codes, their advantages and disadvantages are analyzed, and their comparison with existing codes is carried out.

## II. WAVELET TRANSFORM

In this section, will be explained the idea of wavelet transformation [1-7], more detailed information can be found in the works of Daubechies [3].

Let function  $s(t)$  belong to the Hilbert space  $L^2(R)$  with the scalar product  $\langle f(t), g(t) \rangle = \int f(t)g(t)dt$  and the norm  $\int |s(t)|^2 < \infty$ . The idea of the wavelet transform is based on the partition of the signal  $s(t)$  into two components, approximating  $A_m(t)$  and detailing  $D_m(t)$ .

$$s(t) = A_m(t) + \sum_{i=1}^m D_i(t),$$

where  $m$  denotes the decomposition (reconstruction) level.

In this article, will be used wavelet transformation or rather a spline-wavelet transformation for creating an error detection code. In this paper will be explained the idea of spline-wavelet transformation, only for the splines of the first order, which will be used for the construction of codes.

Let  $X$  be a non-uniform grid of elements,  $X = \{x_j\}_{j \in Z}$ , where  $Z$  is the set of integers. Splines of the first order on the grid  $X$  are defined as follows:

$$\begin{aligned} \omega_j(t) &= (t - x_j)(x_{j+1} - x_j)^{-1}, t \in [x_j, x_{j+1}), \\ \omega_j(t) &= (t - x_j)(x_{j+1} - x_j)^{-1}, t \in [x_{j+1}, x_{j+2}), \\ \omega_j(t) &= 0, t \notin [x_j, x_{j+2}), \end{aligned}$$

where  $\omega_j(t)$  – splines,  $x_j$  – elements of  $X$ .

In the process of wavelet decompositions, some element  $x_k$  is thrown out of the grid  $X$ , after this transformation a new grid  $\tilde{X}$ , on the basis of which new splines  $\tilde{\omega}_j(t)$  are constructed. New and old splines are interconnected. This relationship between the elements  $\tilde{\omega}_j(t)$  and  $\omega_j(t)$  can be shown by the formulas:

$$\begin{aligned} \tilde{x}_j &= x_j, \text{ if } j \leq k - 1, \tilde{x}_j = x_{j+1}, \text{ if } j \geq k, \\ \varepsilon &= x_k, \tilde{\omega}_j(t) = \omega_j(t), \text{ if } j \leq k - 3 \\ \tilde{\omega}_j(t) &= \omega_{j+1}(t), \text{ if } j \geq k \\ \tilde{\omega}_{k-2}(t) &= \omega_{k-2}(t) + \tilde{\omega}_{k-2}(x_k)\omega_{k-1}(t) \\ \tilde{\omega}_{k-1}(t) &= \omega_{k-1}(t) + \tilde{\omega}_{k-1}(x_k)\omega_{k-1}(t) \end{aligned}$$

With the help of spline-wavelet decompositions, it is possible to create a large number of different codes constructions among themselves.

## III. BENT FUNCTION

The measure of nonlinearity is an important characteristic of a Boolean function in cryptography. Linearity and properties close to it often indicate a simple (in a certain sense) structure of this function and, as a rule, represent a rich source of information about many of its other properties. The problem of constructing Boolean

functions possessing nonlinear properties naturally arises in many areas of discrete mathematics. And often the most interesting are those functions for which these properties are extreme. Such Boolean functions are called bent functions. A bent function can be defined as a function that is extremely poorly approximated by affine functions [1].

The nonlinearity of a function  $f$  is the distance from  $f$  to a class of affine functions. Let's denote the nonlinearity of the function  $f$  in terms of  $N_f : N_f = d(f, A(n)) = \min_{g \in A(n)} d(f, g)$ , where  $A(n)$  is the class of linear functions.

The function  $f \in P_2(n)$  is called maximally nonlinear if  $N_f = 2^{n-1} - 2^{(n/2)-1}$ .

*Definition:* A bent function is a Boolean function with an even number of variables for which the Hamming distance from the set of affine Boolean functions with the same number of variables is maximal.

Example:  $f(x_0, x_1, x_2, x_3) = x_0x_1 + x_2x_3$

From the point of view of cryptography, the important criteria that a Boolean function  $f$  of  $n$  variables must satisfy are the following:

1) equilibrium — the function  $f$  takes values 0 and 1 equally often;

2) the propagation criterion  $PC(k)$  of order  $k$ - for any nonzero vector  $y \in Z_2^n$  weight at most  $k$ , the function  $f(x+y) + f(x)$  is balanced;

3) the maximum nonlinearity - the function  $f$  is such that the value of its nonlinearity  $NF$  is maximal;

The bent function matches the criteria for propagation and maximum non-linearity, which allows it to detect all errors in the channel and to have a uniform probability of detecting errors, but the function is not balanced.

#### IV. SPLINE-WAVELET ROBUST CODE

In this section, will be described the rules for the formation of code words for a particular code construction, a comparison of these codes with examples of linear and nonlinear codes are also will be given.

Robust codes are nonlinear systematic error-detecting codes that provide uniform protection against all errors without any (or that minimize) assumptions about the error and fault distributions, capabilities and methods of an attacker [12, 16-18].

Let  $M = |C|$ , this is the number of codewords in code  $C$ . By the definition of an  $R$ -robust code, there are no more than  $R$  code words that cannot be detected for any fixed error  $e$ .

$$R = \max \{ |x| \mid x \in C, x + e \in C \}$$

The probability of masking the error  $e$  can be defined as:

$$Q(e) = \frac{|\{x \mid x \in C, x + e \in C\}|}{M}$$

One of the main criteria for evaluating the effectiveness of a robust code is the maximum error masking probability. The maximum error masking probability can be defined as

$$\max Q(e) = \max \frac{|\{x \mid x \in C, x + e \in C\}|}{M} = \frac{R}{M}$$

The following is the construction of robust codes based on bent functions and spline-wavelets with the static grid,

and grid based on the codeword. The additional elements are calculated on the basis of bent functions from information elements and spline-wavelet elements, the result is also a bent function. The function for the additional elements is the bent function (code Kerdock), the elements are the informational elements and wavelet elements. So, the new function is created, because wavelet elements are the function of several informational elements. This function is also bent function, it was checked for all grid values. The new bent function is created, with another properties.

Let  $c = \{c_1, c_2, \dots, c_{n-1}, c_n\}$  denotes the code word of some shared  $(n, k)$  code. Then  $\{c_1, c_2, \dots, c_{k-1}, c_k\}$  is the information part, and  $\{c_{k+1}, \dots, c_n\}$  - additional.

*Construction 1. Spline-wavelet bent robust code with a static grid.*

In this construction, for all code, a grid is selected  $x = \{x_1, x_2, \dots, x_{n-1}, x_k\}$ , any elements are discarded at the discretion of the specialist, the number of discarded items is equal to  $(n-k)/2$ . Number of characters is strictly even and multiple 4, attitudes  $\frac{k}{n} = \frac{2}{3}$ . The ejected elements will be denoted by the set  $z = \{z_1, \dots, z_{(n-k)/2}\}$ . The wavelet elements will be denoted by the set  $b = \{b_1, \dots, b_{(n-k)/2}\}$ . Let  $c = (c_1, c_2, \dots, c_n)$  - vector field  $GF(2^n)$ ,  $1 \leq i \leq n$ . The vector  $c$  belongs to the code if

$$c_{k+j} = b_j = c_{z_i} + c_{z_i+1} + (x_{z_i+2} + x_{z_i-1})(x_{z_i+2} + x_{z_i})^{-1}(c_{z_i-1} + c_{z_i+1})$$

For even  $z_i$ :

$$c_{k+j+(n-k)/2} = c_1 * c_2 + \dots + b_j * c_{z_i-1} + \dots + c_{k-1} * c_k$$

For odd  $z_i$ :

$$c_{k+j+(n-k)/2} = c_1 * c_2 + \dots + b_j * c_{z_i+1} + \dots + c_{k-1} * c_k$$

Where  $1 \leq j \leq (n-k)/2$ ,  $k$  - the number of parity symbols in the code,  $z_i \in z, +$  - addition mod 2,  $c_{k+j+(n-k)/2}$  is the bent function's element.

This construction is built on a static grid, which is not always good, because it will be necessary to transfer the grid between the receiver and the transmitter.

*Construction 2. Spline-wavelet bent robust code with a grid based on the codeword.*

In this construction, for all code, a grid is selected  $x = \{x_1, x_2, \dots, x_{n-1}, x_k\}$ , based on the information part of the codeword, and depending on the number of the ejected element. The grid is equal to the shift relative to the number of the element that is thrown, that is  $x[i] = c[(i-z_i) \pmod{n-k}]$ . The wavelet elements will be denoted by the set  $b = \{b_1, \dots, b_{(n-k)/2}\}$ . Let  $c = (c_1, c_2, \dots, c_n)$  - vector field  $GF(2^n)$ ,  $1 \leq i \leq n$ . The vector  $c$  belongs to the code if

$$c_{k+j} = b_j = c_{z_i} + c_{z_i+1} + (x_{z_i+2} + x_{z_i-1})(x_{z_i+2} + x_{z_i})^{-1}(c_{z_i-1} + c_{z_i+1})$$

For even  $z_i$ :

$$c_{k+j+(n-k)/2} = c_1 * c_2 + \dots + b_j * c_{z_i-1} + \dots + c_{k-1} * c_k$$

For odd  $z_i$ :

$$c_{k+j+(n-k)/2} = c_1 * c_2 + \dots + b_j * c_{z_i+1} + \dots + c_{k-1} * c_k$$

Where  $1 \leq j \leq (n-k)/2$ ,  $k$  - the number of parity symbols in the code,  $z_i \in z, +$  - addition mod 2.

This construction is built on a grid, based on the codeword, and it solves the problem of the transfer grid, but the algorithm is more time consuming. Created constructions have better parameter than existing, presented example will show the difference between created constructions and existing solutions.

*Example:* Consider the composition of construction 1 and construction 2 for  $n=8$  and  $k=4$ .

In order to obtain the code of this kind, we will consider redundant symbols, as a result of the bent-function of the additional stream  $b$  and other information symbols. It is necessary to make sure that the result of the interaction of the main stream and the additional one is also a bent function. The number of the ejected element is taken equal to three (the number of the ejected element does not affect anything, you can throw out other elements, getting other formulas, but this does not affect the final result).

Formulas for decomposition and reconstruction when the element is kicked out under the number  $k$  have the form, provided that  $x_{k+2} \neq x_k$ :

$$b_k = c_k - c_{k+1} - (x_{k+2} - x_{k-1})(x_{k+2} - x_k)^{-1}(c_{k-1} - c_{k+1})$$

$$c_k = b + c_{k+1} - (x_{k+2} - x_{k-1})(x_{k+2} - x_k)^{-1}(c_{k-1} - c_{k+1})$$

As a bent function for the code, will be take the formula  $f = c_1c_2 + c_3c_4 + c_5c_6 + c_7c_8$ . When the third element will be thrown out. The element  $c_3$  will be replaced on the additional flow element  $b_3$ , and  $c_5$  on the additional flow element  $b_5$ . Functions takes the form  $f_1 = c_1c_2 + b_3c_4 + c_5c_6 + c_7c_8$  and  $f_2 = c_1c_2 + c_3c_4 + b_5c_6 + c_7c_8$ , the addition goes modul 2, corresponds to the operation XOR.

Because these functions are bent functions, so, independently of the values of the grid, the function  $f$  is a bent-function. The result of the function  $f$  will be used as a redundant symbol  $r_0 = f_1$ , and the redundant symbol  $r_1 = b_1$  (element of additional stream),  $r_2 = f_2$ ,  $r_3 = b_2$ .

Let's compare this code with different values of a grid with a linear code and a robust code of the same length. In the example, as the linear code was taken Hamming code (8,4). Redundant symbols for a nonlinear code will be equal to  $r_0 = c_1c_2 + c_3c_4 + c_5c_6 + c_7c_8$ ,  $r_1 = c_1c_3 + c_2c_4 + c_6c_8 + c_5c_7$ ,  $r_2 = c_1c_5 + c_2c_6 + c_3c_7 + c_4c_8$ ,  $r_3 = c_1c_3 + c_2c_4 + c_6c_8 + c_5c_7$  (code Kerdock).

Let's draw up a graph of error detection probabilities for a spline-wavelet code with a static grid — construction 1, for a spline-wavelet code with a codeword-based grid — construction 2, for a “robust” Kerdock code and a linear code, the result is displayed in Figure 1.

The Hamming code does not match the equiprobability of the error, which makes this code vulnerable to attack by the attacker, in contrast to the Kerdock code and construct 1.

The average probability of detecting an error is insignificantly different for the construction 1, the construction 2, and the robust Kerdock code. In the case of

linear code, the probability is uneven, which makes this code vulnerable to attacks on third-party channels.

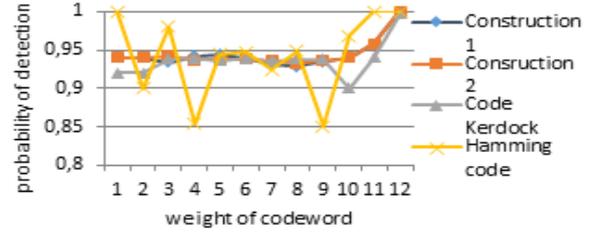


Fig. 1. Error detection probability for different code constructions

Compare the number of the maximum probability of masking errors of these codes and display the results in Table 1, as well as the time, is taken to encode 4000 bytes.

Table 1. Probability of maximum error concealment for different code constructs

Code	maximum probability of error concealment, $Q(e)$	Number of time measurement, sec
Construction 1	0,5	0,045
Robust code Kerdock	0,5	0,067
Hamming code	1	0,037
Construction 2	0,46875	0,067

Codes constructed on the basis of spline-wavelet decompositions using bent functions have a number of advantages in different characteristics compared to other codes. Construction 1 has good coding time, does not have undetectable errors. Construction 2 has the bad coding time, but has no undetected errors, and has the smallest value of the maximum probability of masking the error.

For codes based on bent functions and wavelet transformations, there will always be an even number of information symbols, since bent functions exist only for an even number of variables, and codes have poor coding time scores compared to linear codes.

## V. SPLINE-WAVELET CODE ON BENT FUNCTION WITH DIFFERENT DEGREES

In the case of construction 1, the degree of the bent functions is equal to 2, in the case of the construction 2, the degree of the bent functions is 3 or 2, depending on the symbols to be erased. As a result, an assumption immediately arises that an increase in the degree of the curved function can give the best values for the parameter  $R$ .

The degree of the bent function cannot exceed  $n/2$  [8], and therefore one can not single out a single construction for an arbitrary number of variables.

In this section will be shown functions with different degrees based on spline wavelets and information symbols for  $n = 8$ , they will be compared by the parameter  $R$ . For each number of variables, bent functions were constructed on the basis of spline wavelets. For all functions, the wavelet element is calculated from the same function:

$$Wave_k = c_k - c_{k+1} - (x_{k+2} - x_{k-1})(x_{k+2} - x_k)^{-1}(c_{k-1} - c_{k+1})$$

Functions for n=8 presented in table 2, also given the conditions of the grid and the degree of function.

Table 2. Spline wavelet Bent functions for n=8

Number of function	Grid	Function	Deg(f)
1	$x_i = c_i$	$f_i = c_{i+1} * c_{i+3} * c_{i+4} + c_{i+2} * c_{i+3} * c_{i+5} + Wave_{i+2} * c_{i+6} + c_i * c_{i+3} + c_i * c_{i+5} + c_{i+2} * c_{i+3} + c_{i+2} * c_{i+4} + c_{i+2} * c_{i+5} + c_{i+3} * c_{i+4} + c_{i+3} * c_{i+5} + c_{i+6} * c_{i+7}$	4
2	Static	$f_i = c_i * c_{i+1} * Wave_{i+2} + c_{i+1} * c_{i+3} * Wave_{i+4} + c_i * c_{i+1} + c_i * c_{i+3} + c_{i+1} * c_{i+5} + c_{i+2} * c_{i+4} + c_{i+3} * c_{i+4} + c_{i+6} * c_{i+7}$	3
3	$x_i = c_{7-i}$	$f_i = c_i * c_{i+1} + Wave_{i+2} * c_{i+3} + c_{i+4} * c_{i+5} + c_{i+6} * c_{i+7}$	4
4	Static	$f_i = c_i * Wave_i + c_{i+1} * Wave_{i+2} + c_{i+2} * Wave_{i+4} + c_i * c_{i+3} + c_{i+1} * c_{i+5} + c_{i+2} * c_{i+3} + c_{i+2} * c_{i+4} + c_{i+2} * c_{i+5} + c_{i+3} * c_{i+4} + c_{i+3} * c_{i+5} + c_{i+6} * c_{i+7}$	2

Let's compile the code constructions for all the above functions with 2 redundant symbols  $r_0 = f_0, r_1 = f_1$ . Calculate the parameter R and maximum probability of error concealment, the results are listed in Table 3.

Table 3. Parameter R for n=8

Function	The degree of bent function	R	maximum probability of error concealment, Q(e)
Function №1	4	96	0,375
Function №2	3	120	0,46875
Function №3	4	96	0,375
Function №4	2	128	0,5

The degree of the bent function different from 2 gives a better result for the parameter R. The number of calculations and the time spent on coding information is more compared to the codes built on bent functions with a power of 2. When these codes are used in the case protection against attack by an attacker, then the parameter R is more important. Using spline wavelets, it is possible to build a large number of robust codes, build bent functions and increase their degree, thereby improving the quality of robust codes. So new construction is created with parameter R lower than construction 1 or construction 2.

*Spline-wavelet robust code with lower value R*

Let  $c = \{c_1, c_2, \dots, c_{n-1}, c_n\}$  denotes the code word of some shared  $(n, k)$  code. Then  $\{c_1, c_2, \dots, c_{k-1}, c_k\}$  is the information part, and  $\{c_{k+1}, \dots, c_n\}$  - additional,  $n = k + 2$ . Grid is selected depending on the spline wavelet function,  $f_i(c_1, c_2, \dots, c_{k-1}, c_m)$  is a function from table 2,  $m = 8$ . The vector  $c$  belongs to the code if

$$c_{k+1} = f_0(c_1, \dots, c_m) + c_{m+1} * c_{m+2} + \dots + c_{k-1} * c_k;$$

$$c_{k+2} = f_1(c_1, \dots, c_m) + c_{m+1} * c_{m+2} + \dots + c_{k-1} * c_k.$$

This construction allows better protection against side-channel attacks, because parameter R and maximum probability of error concealment Q(e) lower than existed solutions, but it takes more time for the coding information.

**Conclusion**

In this paper, was described as the error-correcting coding scheme based on wavelet transformation and bent functions. For the proposed scheme, was created wavelet robust codes on bent functions. The robust wavelet code has no undetectable errors, so it ensures reliable protection against the error injection, also has the high values of the parameter R.

**REFERENCES**

- [1] I. G. Burova and U. K. Demyanovich, Theory of Minimal Spline (SPSU, 2000).
- [2] G. Caire, R. L. Grossman and H. V. Poor, Wavelet transforms associated with finite cyclic groups, IEEE Trans. Inf. Theory 39(4) (1993) 1157–1166.
- [3] I. Daubechies, Ten Lectures on Wavelets, CBMS-NSF Conference Series in Applied Mathematics (SIAM, 1992).
- [4] U. K. Demyanovich, Calibration ratio for B-splines on nonuniform net, Mat. Model T 13(3) (2001)
- [5] U. K. Demyanovich, Minimal Splines and Wavelets (Vestnik SPSU, 2008)
- [6] F. Fekri, R. M. Mersereau and R. W. Schafer, Theory of wavelet transform over finite fields, IEEE International Conference on Acoustics, Speech, and Signal Processing 3 (1999) 1213–1216.
- [7] F. Fekri, S. W. McLaughlin, R. M. Mersereau and R. W. Schafer, Double circulant selfdual codes using finite-field wavelet transforms, Applied Algebra, Algebraic Algorithms and Error Correcting Codes Conference (Springer, 1999), pp. 355–364
- [8] Tokareva N., Bent Functions: Results and Applications to Cryptography, 2015.
- [9] A. Levina and S. Taranov, Spline-wavelet robust code under nonuniform codeword distribution, in 3rd Int. IEEE Computer, Communication, Control and Information Technology (IEEE, 2015).
- [10] A. B. Levina and S. V. Taranov, Algorithms of constructing linear and robust codes based on wavelet decomposition and its application, Cryptology, and Information Security (Springer, 2015), pp. 247–258.
- [11] A. B. Levina and S. V. Taranov, Second-order spline-wavelet robust code under nonuniform codeword distribution, Procedia Comput. Sci. 62 (2015) 297–302.
- [12] A. B. Levina and S. V. Taranov, Construction of linear and robust codes that is based on the scaling function coefficients of wavelet transforms, J. Appl. Ind. Math. 9(4) (2015) 540–546.
- [13] A. B. Levina and S. V. Taranov, New construction of algebraic manipulation detection codes based on wavelet transform, Proceedings of the 18th Conference of Open Innovations Association FRUCT - 2016, pp. 187-192
- [14] A. B. Levina and S. V. Taranov, Creation of codes based on wavelet transformation and its application in ADV612 chips, International Journal of Wavelets, Multiresolution and Information Processing - 2017, Vol. 15, No. 2, pp. 1750014
- [15] A. B. Levina and S. V. Taranov, AMD codes based on wavelet transform, 2017 Progress In Electromagnetics Research Symposium - Fall (PIERS-FALL) - 2017, pp. 2534-2539
- [16] Akdemir K.D., Wang Z., Karpovsky M. G., Sunar B., Design of Cryptographic Devices Resilient to Fault Injection Attacks Using Nonlinear Robust Codes // Fault Analysis in Cryptography, 2011.
- [17] Karpovsky M.G., Kulikowski K., Wang Z., Robust Error Detection in Communication and Computation Channels // Keynote paper, Int. Workshop on Spectral Techniques, 2007.
- [18] Carlet C. Boolean functions for cryptography and error correcting codes // Chapter of the monograph «Boolean Methods and Models», Cambridge Univ. Press (P. Hammer, Y. Crama eds.), 2007.
- [19] MacWilliams, F.J. and Sloane, N.J.A., The Theory of Error-Correcting Codes. Elsevier-North-Holland, Amsterdam, 1977.

# Cryptographic keys management system based on DNA strands

Marek Miśkiewicz  
Maria Curie-Skłodowska University  
Plac Marii Curie-Skłodowskiej 5  
20-031 Lublin, Poland  
Email: marek.miskiewicz@umcs.pl

Bogdan Księżopolski  
Maria Curie-Skłodowska University  
Plac Marii Curie-Skłodowskiej 5  
20-031 Lublin, Poland  
Email: bogdan.ksiezopolski@umcs.pl

**Abstract**—Security of cryptographic keys is one of the most important issues in a key management process. The question arises whether modern technology really allows for a high level of physical protection and security of sensitive data and cryptographic keys. The article considers various contemporary types of threats associated with the storage of secret keys. We present an innovative way to store sensitive data, using DNA strands as a medium, which significantly reduces hazard connected with electronic devices based data storage and makes the key management process independent of third parties.

## I. INTRODUCTION

SECRET keys are usually stored in hard drives placed in computer devices with access secured with a simpler several-character password or on portable flash drives. Such solutions have many disadvantages and really do not guarantee a high level of data security or even in some cases usability. There are a few important reasons why the cryptographic data is not completely safe if it is stored on portable devices (in NAND memory chips) or magnetic storage devices. It is obvious that access to stored cryptographic keys needs at least computer devices with access to wider resources eg. Internet network. With the current complexity of digital systems, we can not fully guarantee security, which to some extent, relies on trust in the integrity of digital system manufacturers and designers. Absolute security, at least theoretically, can only be guaranteed by the lack of participation of third parties in the process of managing and storage of cryptographic keys (creation, storage, use and destruction). If we consider it necessary to use electronic devices, this entails additional risks, in particular, due to the lack of access to stored data during power system failure caused for example by Solar Storm. In the article, the new cryptographic keys management system based on DNA strands is presented. The contributions of our concept are as follows:

- security increase by excluding third parties from the process of storing and reading the key,
- lack of weaknesses and vulnerabilities associated with storing cryptographic keys on portable electronic devices,
- no susceptibility to data destruction caused by strong electromagnetic fields,
- enormous difficulty in accessing data (e.g. cryptographic keys) by unauthorised persons in case of loss of control,

- no need to use DNA sequencing devices to read stored data (minimal resources infrastructure),
- data transferability in a way that can be almost completely undetectable in physical form.

## II. RELATED WORK

In the domain of cryptography key management systems based on biosystems one can analyse three issues: bio cryptography, third-party key management and environmental threats.

### A. Bio cryptography

The use of DNA strands to store information and even perform simple "calculations" is not a completely new idea. Adelman in his work showed the possibility of using fragments of specially prepared DNA strands to solve the problem of Hamilton's path [1]. Gehani together with others created the basis of DNA-based cryptosystem based on the idea of One Time Pad [2]. In the work of Y. Zhang, X. Lui and M. Sun a practical implementation of the problem of key distribution for the OTP method was shown [3]. The sequence of nucleotides in a randomly selected fragment of DNA is used as the key to encrypt the message. The explicit text has been replaced with a sequence of bits and using the XOR function joined with the key string. The key, based on the DNA sequence can be obtained by using one of the possible substitutions of nucleotides: A - 00, C - 01, T - 11, G - 10. Next, the "DNA key" was „glued" to the plasmid and placed in the bacterial cell. The environment inside the bacteria allows you to stably hold the information contained in the DNA strand, which is very sensitive to changes in the temperature and pH of the solution in which it is located. The stability of the DNA accumulated in bacterial cells carried out in the state of spore is impressive. Scientists have been able to read genetic material from *Subtilis* bacteria, which is millions of years old [4], [5]. Modern laboratory techniques allow for stable storage synthetic DNA in Silica for thousands of years [6], [7]. This may be important if it is necessary to store relevant information (in particular cryptographic information) for a very long period of time. Traditional storage technologies such as magnetic devices and optical discs are not reliable for really long-term data storage. Their lifespan is estimated to be about 50 years [8].

Halvorsen and Wong in his paper [9] showed an interesting, simple and secure system for encrypting and decrypting information using self-assembly DNA structures and PCR based decoded information reading method. Tanaka, Okamoto and Saito presented a system for public key distribution based on DNA as a one-way function [10]. Using the methods and algorithms described in the works of A. Leier [11] and H. J. Shiu [12], one can hide the message in a DNA sequence in an encrypted or unencrypted way. Such stenographic techniques require active synthesis of deoxyribonucleic acid chains. The text is encrypted directly in the series of A, C, T and G, or special groups are identified later as counterparts of binary zeros and ones. The presented methods require both synthesis and sequencing devices at almost every stage of work with data stored in DNA, which seems to be an inconvenience in a certain class of applications. Some ideas presented in the last two publications are applied further. The DNA chain can also be successfully used in forensics [13] as well as for invisible product tagging [14].

### B. Third-parties keys management

A user who really cares about the security of his or her data cannot be sure that the data storage devices, produced by third parties, do guarantee real security. This is due to the fact that the average user does not have access to the exact device specification and is not able to check if the electronic systems controlling the memory chips do not allow easy access to data stored by unauthorized entities. In other words, strong cryptography and ultimate cryptographic keys security require the assumption of complete distrust in the devices that are used. The continuous reduction in the size of integrated circuits leads to increased production costs. This forces a vast majority of chip design companies to trust an external third party in chip fabrication, but outsourcing of chip fabrication opens-up hardware to attack. The way of preparing post fabrication tests leave an open door for implementing malicious modifications and backdoors. Even if there are no equipment manufacturers bad intentions there is always a possibility that there has been interference of third parties.

Researchers at the University of Michigan showed in their paper [15], that there is a possibility to create a novel fabrication-time attack based on modifications of the semiconductor structure in integrated circuits. It can be done by adding even single component to "mask" - a blueprint of the chip before its production. Such modifications are hardly detected during the test procedure. This kind of attack is triggered by special *unlike* sequence of commands and allows to give a malicious program the full operating system access.

Other, pernicious fabrication-time attack named dopant-level Trojan bases on conversion trusted circuits into malicious circuitry in chip structure by changing the dopant ratio on the input pins to transistors [16], [17]. Circuits converted to Trojans are very difficult to detect due to the lack of added circuit elements and require imaging with a scanning electron microscope.

Spiegel Online reports in his article [18] that the US National Security Agency (NSA) is in possession of specially prepared "computer buggung" devices that look like typical USB plugs. These devices are capable of sending and receiving data via radio link being undetected.

In October 2018 Bloomberg reported that special microchips were inserted into server motherboards during the production process. The motherboards are components of servers operating in many companies inside their datres. Some of these chips were built as if they were necessary elements for the proper operation of the entire system. Installed chips have enough processing power to carry out an attack or be used to gain unauthorized access to data [19].

### C. Environmental threats

One of the important factors that should be taken into account when cryptographic data is stored on electronic devices and magnetic storage devices is their relatively high sensitivity to strong electromagnetic fields. These kinds of fields can be produced in two ways: as EMP pulses (Electro-Magnetic Pulse) or during the Electromagnetic Solar Storm, especially in the so-called Coronal Mass Ejection.

It is worth to mention at least about two cases of CME, which had a significant impact on the human created infrastructure. The first one is The Solar Storm of 1859 (known as Carrington Effect). During this storm, Earth's magnetic field disturbances caused by CME led to telegraph network failures throughout Europe and North America in some cases giving telegraph operators electric shocks. The second one took place on March 13, 1989. A severe geomagnetic storm struck Earth causing nine hours blackout in Quebec, Canada.

Report prepared by Metatech Corporation [20] describes the threat of the early-time (E1) High-altitude Electromagnetic Pulse (HEMP) produced by nuclear detonations above an altitude of 30 km. The pulse is driven by gamma photons produced in nuclear reactions within the nuclear burst. The main impact has such impulse on the power grid that can be totally damaged, but as the report shows, computers and small electronic devices are also at risk of damage what can make them unusable.

## III. THE METHOD

In this paper, we use and extend the concept presented in the work of [11], where the single bits of information are represented by groups of nucleotides. With certain restrictions, this solution allows to easily generate sequences of data stored in DNA without the need to use synthesis devices. If the prepared data contain secret information, for example, password or cryptographic keys then, as it was mentioned earlier, the lack of third parties engagement during the synthesis process significantly increases data security.

### A. DNA data structure

DNA strand with data stored within, consist of a series of specially prepared components – some kind of building blocks which in fact are shorter fragments of double-stranded DNA

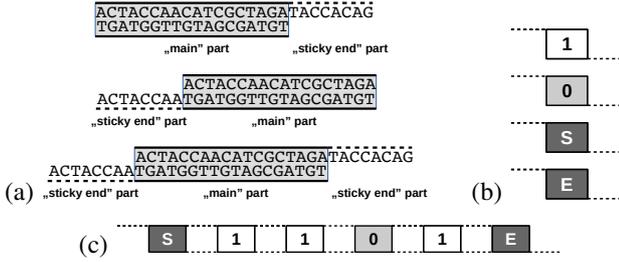


Fig. 1. (a) Structure of simple DNA component. Structural components to build simple DNA data strand. (b) Structural components to build simple DNA data strand. (c) Example of structure of four bit DNA strand.

ended with single-stranded regions called sticky ends. The general shape of a single components is presented in Figure 1 (a). The main part identifies some sort of data stored as a series of nucleotides. These nucleotides can represent a bit or index of extracting parts. Sticky ends are free fragments of one-sided DNA strand that can easily bind to other complementary part connected to the other fragments to producing longer strands.

In the simplest case to store data in DNA as a series of 0 and 1 bits one needs at least four structurally different fragments (see Figure 1 (b)). Two of them named  $S$  and  $E$  starts and ends DNA strand containing data. "0" and "1" fragments represent bis of data. An example of general structure of four bits single DNA strand is shown in the Figure 1 (c).

Every DNA strand containing data bits always starts with  $n$ -numbered  $S$  fragment. Start fragment ( $S_n$ ) is a double-stranded fragment of DNA with length about 30 bp (nucleobase pairs). It consists of the "main" part and sticky end part. The main part (approximately 22 bp) carries information that can be used to identifying bit sequence as a part of a larger amount of data. It can be also used as a primer for the PCR procedure. Sticky end part (length about 8 bp - depended on the total length of the whole strand - explained later in the text) allows binding with next structure fragment - bit fragment.

Bit fragment ( $B_{0k}$  or  $B_{1k}$ ) is a double-stranded fragment of DNA of length about 20 - 30 bp. It consists of Bit identification part and two sticky end parts. There are two different types of Bit identification part - one for bit "1" and the other for bit "0". In the simplest approach the internal structure of every bit "1" and "0" are the same for the whole data strand. In this case, there is no need to use sequencing devices to read data from DNA. Data sequence from the specified strand can be retrieved only by gel electrophoresis.

End fragment is about 20 bp length and it is the last structure element of data DNA strand. Like other fragments, it contains an identification part and a sticky end part.

**B. Sticky ends**

The sticky end is a short single-stranded fragment of DNA placed on its end that allows binding DNA fragments into longer strands. The Special design of the sticky ends nucleotide sequence allows the complementary fragments to join generating a fixed order of components. The idea of how the sticky ends work is presented on Figure 2.

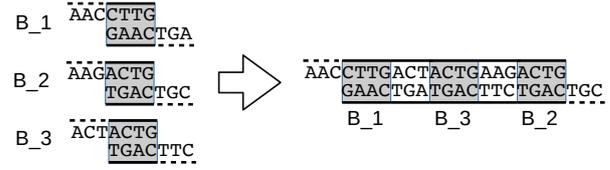


Fig. 2. Shorts three fragments of DNA with complementary sticky ends before ligation.

$N$  bits data string encoded into DNA strand requires at least:

- DNA start component ( $S$ ) with sticky end marked as  $s_0$ ,
- $N$  pair of "Bit" components:
  - first pair of bits with structure:  $s_0 - 0 - s_1$  and  $s_0 - 1 - s_1$ , where  $s_n$  means  $n$ 'th sticky end,
  - second pair of bits with structure:  $s_1 - 0 - s_2$ ,  $s_1 - 1 - s_2$ ,
  - $n$ 'th bits pair structure:  $s_{n-1} - 0 - s_n$ ,  $s_{n-1} - 1 - s_n$ ,
- end component ( $E$ ) with sticky end marked as  $s_n$ .

DNA data strand structure looks like this:  $S - s_0 - B_1 - s_1 - B_2 - s_2 - \dots - s_{n-1} - B_n - s_n - E$ .  $B_n$  denotes "0" or "1" bit component. As one can see  $n$  different sticky ends  $s_i$  are required. Due to the fact that sticky ends consist of  $k$  nucleotides, only  $4k$  different nucleotide sets can be produced. From the statistical point of view, for data strand containing  $n$  bits minimal length of each sticky end should be at least:  $k = \frac{1}{2} \log_2 n$ . For example, 64-bit data strand requires sticky ends that consist of only 3 nucleotides. In fact, as it could be seen in [11], even for 8-bit data strands structure of individual DNA components is more complicated and requires sticky ends of length 10 nt (nucleotides). This is due to the fact that biochemical conditions and processes play a significant role in the problem of sticky end creation and use. The simplest case of minimal required sticky ends length is insufficient and do not lead to the successful creation of longer data strands. Biological limitations related to the procedure of creation data stands from DNA fragment and read them by gel electrophoresis cause that the number of bits carried by DNA strand is not enough to store for example long cryptographic key (1024 to 4096 bits) in a single strand. The reasonable total length of DNA strand that can be used for data storage considered in this paper is about 1000 bp. For such strands number of stored bits is about 32, so 1024 bit keys require 32 different DNA strands. It is, therefore, necessary to introduce a system of indexing individual strands or even individual keys if the multi-key system is introduced.

**C. Single  $n$ -length key DNA data strand preparation**

Let us consider the user that wants to store  $n = 2^k$  bits long cryptographic key in the DNA strand, that can carry only  $m = 2^{k-l}$  bits. Thus he requires  $2^l$  DNA strands, where every one of them contains  $m$  bits of key called subkey. Structure of a single DNA strand from the given set is as follows:  $s_i - S_i - s_0 - B_1 - s_1 - B_2 - s_2 - \dots - s_{m-1} - S_{m-1} - s_m - B - s_m - E_j - s_j$ .  $i$  and  $j$  denote subkey number and vary from 1 to  $2^l$ . As one can see  $2^l$   $S$  and  $E$  fragments must be synthesised with different unique

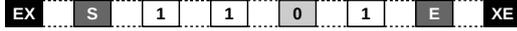


Fig. 3. Example of 4 bit data strand with Extractor fragments bounded at the ends.

sticky ends. The purpose of that will be explained further. Finally, the procedure of preparation DNA strands for store  $n$  bit length key has following steps:

- 1) Generation and synthesis of  $2^l$  unique DNA Start fragments  $S$ :  $s_i - S_i - s_0$ .
- 2) Generation and synthesis of  $m$  pairs of Bit fragments  $s_0 - 0_1 - s_1$  to  $s_{m-1} - 0_m - s_m$  and  $s_0 - 1_1 - s_1$  to  $s_{m-1} - 1_m - s_m$ . All Bit fragments are called Bit Library.
- 3) Generation and synthesis of  $2^l$  unique DNA End fragments  $E$ :  $s_m - E_i - s_j$ .
- 4) Preparation of subkeys  $K_{m:i}$  for  $i = 1$  to  $2^l$  by splitting key  $K$  into  $2^l$  fragments. Subkey  $K_{m:i} = \{B_1, B_2, \dots, B_m\}_{m:i}$  where  $B_1, B_2, \dots$  represent individual bits of subkey.
- 5) For the first subkey of key  $K$  mix in the reaction tube Start fragments  $S_1$ , End fragments  $E_1$ , and a set of Bit fragments chosen from Bit Library in a way to match the corresponding first subkey bits. Then incubate mixture according to biotechnological protocols to obtain double-stranded DNA.
- 6) Repeat previous step for next subkey of key  $K$ .

After a procedure of generating key  $K$  one should have  $2^l$  reaction tubes with subkeys. Content of reaction tubes can be mixed together in one tube after DNA purification. The final tube contain all the key  $K$  stored in DNA as a series of its bits. Presented procedure can be extended to store more than one key  $K$  just in one tube. Comments require the presence of sticky ends denoted as  $s_i$  and  $s_j$  at the beginning and at the end of the strand. This is straightly connected with subkey extraction and the read procedure described below.

#### D. Subkey extraction

To read a sequence of bits of key  $K$  a sequence of each subkey must be read. All subkey are stored in one tube so the procedure of extraction single subkey must exist. This can be done by preparing a special set of extractors called  $EX$  and  $XE$  which are in fact fragments of double-stranded DNA ended by sticky ends at one side. The extractors are designed to bind to a selected strand representing subkey both at the start ( $EX$ ) and the end ( $XE$ ). It provides to extend the length of the subkey strand as it is shown on a Figure 3.

There is a pair of primers designed to be complementary to extractor pair. Primers are needed for the PCR procedure to increase the number of DNA strands carried extracted subkey. Primers are called  $PEX$  and  $PXE$ . The general procedure to extract subkey  $K_{m:i}$  is as follows: i. get a small amount of mixture containing the key  $K$  and put it into another reaction tube, ii. add pair of extractors  $EX_i$  and  $XE_i$  to the reaction tube and ligate them, iii. prepare and proceed electrophoresis on an agarose gel to separate strands

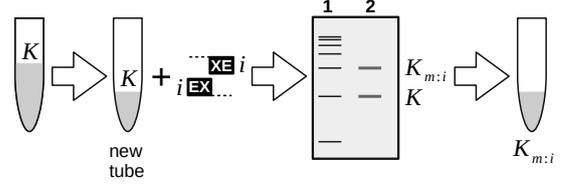


Fig. 4. Schematic process of subkey  $K_{m:i}$  extraction.

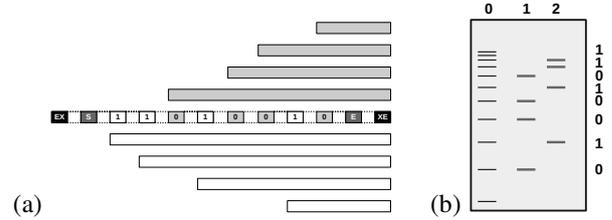


Fig. 5. (a) Expected primer length after elongation in case of 8 bit encoding DNA strand after PCR procedure. (b) An example picture of gel electrophoresis performed for strands set on the left side. Lane 0 symbolize distribution of molecular weight marker, lane 1: distribution of strands length elongated with "0" primer, lane 2: distribution of strands length elongated with primer "1". Reading from bottom to top reveals an encoded bit sequence.

extended by the extractors, vi. isolate from gel DNA strands, that the length corresponds to the length of subkey DNA plus the length of both extractors, v. for extracted DNA material provide the PCR process with primers  $PEX_i$  and  $PXE_i$  relevant to extractors  $EX_i$  and  $XE_i$  to amplify a number of copies of DNA strand.

#### E. Subkey reading

To determine a series of bits in extracted and isolated subkey one must perform two-step procedure used by [11] in his work. The first step is to carry out PCR procedure with two types of primers. Solution with isolated and replicated subkey must be split into two reactions tubes. To the first tube one has to add primers corresponding to "0" bit DNA fragment, to the second reaction tube primers corresponding to "1" bit fragment must be added. Next for both tubes PCR must be performed to elongate the primers. After PCR reaction tubes should contain shorter DNA strands with length matching to the position of "0" and "1" fragments. Figure 5 a. shows example of PCR performed for strand encoding 8 bit sequence: 1 1 0 1 0 0 1 0. The second step requires the implementation of gel electrophoresis for PCR'ed mixture with a subkey. Contents of both reaction tubes must be put into gel separately on different lanes to visualize "0" bit bands and "1" bit bands. Positions of each band are related to DNA strand length in the analysed sample. Due to the fact that Bit fragments forming subkey consist of a determined number of nucleotides, some kind of "quantisation" must occur after electrophoresis. In other way bands on the gel always should come up at the fixed positions indicating positions of zeroes and ones in the analysed subkey. Picture 5 (b) shows expected bands distribution for example from picture 5 (a). To read a sequence of entire key  $K$  every subkey must be read in mentioned way.

### F. Molecular keyring

A tube containing many keys (stored in DNA) could be considered to be "a molecular keyring". The idea of storing multiple keys in DNA bands is not very different from the method of storing a single key, that can be expanded relatively easy. Such approach requires the creation of a revocation key mechanism and an extractors database for keys identification. As it was mentioned earlier, information about the key number and its subkeys is stored in the first segment of each strand containing sticky ends. A user of this system needs to know which extractors use to obtain a chosen key (i.e. subkeys belonging to this key), so the external information binding extractors (with specified sticky ends) with key structure (subkeys sequence) must exist. This could be done for example by signing tubes containing extractors with a signature like this:  $K_{k:m:i}$  that means: extractor for  $i$ -th subkey of length  $m$  of key  $k$ .

For storing, extracting and reading four 1024 bit keys as a series of many 32 bit sequences (which is, in fact, one 4096 bit key - mostly use in e.g. RSA system) user needs in total less than 512 tubes of oligonucleotides to perform simple operations.

A simple revocation mechanism can be proposed for keys that were used and are no longer valid. Other teys for further processing (e.g. reading) need to be extracted by binding them with  $EX$  and  $XE$  extractors. The sticky ends of  $S$  and  $E$  fragments can be blocked against using them as binding sites by adding to the main tube containing keys short single-stranded DNA fragments called caps. Caps are complementary to the sticky ends that need to be blocked. After ligation using ligase enzyme sticky ends at both ends of selected strands (subkeys) should become inactive and no longer can be used for the key extraction. Revocation system for keyring from the above example needs to manage additional  $4 * 32 * 2 = 256$  tubes of oligonucleotides.

### IV. CONCLUSIONS AND FUTURE WORK

A simple cryptographic keys creation and management system based on DNA strands was presented. Despite the considerable complexity due to the relatively large number of necessary elements the system does not require the participation of third parties in very important steps such as the key creation and reading. These features make it resistant to attacks of stealing data through untrusted (unsecured) elements of IT infrastructure or through access by unauthorized entities. In addition, data stored as molecular structures are not susceptible to EMP or SolarStorms and are largely independent of power grids. The next step should focus on experimental verification of the whole process.

### REFERENCES

- [1] L. M. Adleman, "Molecular computation of solutions to combinatorial problems." *Science*, vol. 266, no. 5187, pp. 1021–1024, Nov 1994. doi: 10.1126/science.7973651. [Online]. Available: <https://doi.org/10.1126/science.7973651>
- [2] A. Gehani, T. LaBean, and J. Reif, *DNA-based Cryptography*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 167–188. ISBN 978-3-540-24635-0. [Online]. Available: [https://doi.org/10.1007/978-3-540-24635-0\\_12](https://doi.org/10.1007/978-3-540-24635-0_12)
- [3] Y. Zhang, X. Liu, and M. Sun, "Dna based random key generation and management for otp encryption." *Biosystems*, vol. 159, pp. 51–63, Sep 2017. doi: 10.1016/j.biosystems.2017.07.002. [Online]. Available: <https://doi.org/10.1016/j.biosystems.2017.07.002>
- [4] R. Cano and M. Borucki, "Revival and identification of bacterial spores in 25- to 40-million-year-old dominican amber," *Science*, vol. 268, no. 5213, pp. 1060–1064, 1995. doi: 10.1126/science.7538699 Cited By 348. [Online]. Available: <https://doi.org/10.1126/science.7538699>
- [5] R. Vreeland, W. Rosenzweig, and D. Powers, "Isolation of a 250 million-year-old halotolerant bacterium from a primary salt crystal," *Nature*, vol. 407, no. 6806, pp. 897–900, 2000. doi: 10.1038/35038060 Cited By 414. [Online]. Available: <https://doi.org/10.1038/35038060>
- [6] R. N. Grass, R. Heckel, M. Puddu, D. Paunescu, and W. J. Stark, "Robust chemical preservation of digital information on dna in silica with error-correcting codes," *Angewandte Chemie International Edition*, vol. 54, no. 8, pp. 2552–2555, 2015. doi: 10.1002/anie.201411378. [Online]. Available: <https://doi.org/10.1002/anie.201411378>
- [7] J. P. Cox, "Long-term data storage in dna." *Trends in Biotechnology*, vol. 19, no. 7, pp. 247 – 250, 2001. doi: 10.1016/S0167-7799(01)01671-7. [Online]. Available: [https://doi.org/10.1016/S0167-7799\(01\)01671-7](https://doi.org/10.1016/S0167-7799(01)01671-7)
- [8] S. B. Shah and J. G. Elerath, "Reliability analysis of disk drive failure mechanisms," *Annual Reliability and Maintainability Symposium, 2005. Proceedings.*, pp. 226–231, 2005. doi: 10.1109/RAMS.2005.1408366. [Online]. Available: <http://doi.org/10.1109/RAMS.2005.1408366>
- [9] K. Halvorsen and W. P. Wong, "Binary dna nanostructures for data encryption," *PLOS ONE*, vol. 7, no. 9, pp. 1–4, 09 2012. doi: 10.1371/journal.pone.0044212. [Online]. Available: <https://doi.org/10.1371/journal.pone.0044212>
- [10] K. Tanaka, A. Okamoto, and I. Saito, "Public-key system using dna as a one-way function for key distribution," *Biosystems*, vol. 81, no. 1, pp. 25 – 29, 2005. doi: 10.1016/j.biosystems.2005.01.004. [Online]. Available: <https://doi.org/10.1016/j.biosystems.2005.01.004>
- [11] A. Leier, C. Richter, W. Banzhaf, and H. Rauhe, "Cryptography with dna binary strands." *Biosystems*, vol. 57, no. 1, pp. 13–22, Jun 2000. doi: 10.1016/S0303-2647(00)00083-6. [Online]. Available: [https://doi.org/10.1016/S0303-2647\(00\)00083-6](https://doi.org/10.1016/S0303-2647(00)00083-6)
- [12] H. Shiu, K. Ng, J. Fang, R. Lee, and C. Huang, "Data hiding methods based upon dna sequences." *Information Sciences*, vol. 180, no. 11, pp. 2196 – 2208, 2010. doi: 10.1016/j.ins.2010.01.030. [Online]. Available: <https://doi.org/10.1016/j.ins.2010.01.030>
- [13] J.-M. Oh, D.-H. Park, and J.-H. Choy, "Integrated bio-inorganic hybrid systems for nano-forensics." *Chem. Soc. Rev.*, vol. 40, pp. 583–595, 2011. doi: 10.1039/C0CS00051E. [Online]. Available: <http://dx.doi.org/10.1039/C0CS00051E>
- [14] S. Cormier, J. Shearman, and M. Hogan, "Dna in your jeans? effect of abrasion and bleaching on dna tagged denim." *AATCC Review*, vol. 18, pp. 44–48, 09 2018. doi: 10.14504/ar.18.5.4. [Online]. Available: <https://doi.org/10.14504/ar.18.5.4>
- [15] K. Yang, M. Hicks, Q. Dong, T. Austin, and D. Sylvester, "A2: Analog malicious hardware," in *2016 IEEE Symposium on Security and Privacy (SP)*, May 2016. doi: 10.1109/SP.2016.10. ISSN 2375-1207 pp. 18–37. [Online]. Available: <https://doi.org/10.1109/SP.2016.10>
- [16] G. T. Becker, F. Regazzoni, C. Paar, and W. P. Burleson, "Stealthy dopant-level hardware trojans: extended version," *Journal of Cryptographic Engineering*, vol. 4, no. 1, pp. 19–31, Apr 2014. doi: 10.1007/s13389-013-0068-0. [Online]. Available: <https://doi.org/10.1007/s13389-013-0068-0>
- [17] R. Kumar, P. Jovanovic, W. Burleson, and I. Polian, "Parametric trojans for fault-injection attacks on cryptographic hardware," in *2014 Workshop on Fault Diagnosis and Tolerance in Cryptography*, Sep. 2014. doi: 10.1109/FDTC.2014.12 pp. 18–28. [Online]. Available: <https://doi.org/10.1109/FDTC.2014.12>
- [18] J. Appelbaum, J. Horchert, O. Reissmann, M. Rosenbach, J. Schindler, and C. Stöcker. (2013, 12) Unit offers spy gadgets for every need. [Online]. Available: <http://www.spiegel.de>
- [19] J. Robertson and M. Riley. (2018, 11) The big hack: How china used a tiny chip to infiltrate u.s. companies. [Online]. Available: <https://www.bloomberg.com>
- [20] E. Savage, J. Gilbert, and W. Radasky, "The early-time (e1) high-altitude electromagnetic pulse (hemp) and its impact on the u.s. power grid," Metatech Corporation, 358 S. Fairview Ave., Suite E Goleta, CA 93117, Tech. Rep., January 2010.



# Malicious and Harmless Software in the Domain of System Utilities

Jana Šťastná

Technical university of Košice  
Letná 9, 042 00 Košice, Slovakia  
Email: jana.stastna@tuke.sk

**Abstract**—The focus of malware research is often directed on behaviour and features of malicious samples that stand out the most. However, our previous research led us to see that some features typical for malware may occur in harmless software as well. That finding guided us to direct more attention towards harmless samples and more detailed comparisons of malware and harmless software properties. To eliminate variables that may influence the results, we narrowed down our research study to specific software domain - system maintenance and utility tools. We analysed 100 malicious and 100 harmless samples from this domain and statistically evaluated how they differ regarding packing, program sections and their entropies, amount of code outside common sections and we also looked at differences in behaviour from the high-level view.

## I. INTRODUCTION

WHEN studying research papers in the field of malware research, one may get the impression that harmless software is somehow neglected in research studies. The idea we have specifically in mind is that presence or absence of specific features or differences in their qualities in malware with comparison to harmless software is seldom examined. However, such studies would be of great help.

Many research works use harmless software only as a resource for demonstrating detection rate of new presented detection method. However, selection of harmless samples may considerably influence detection results and thus detection rates. When programs that are part of the default system installation are used as a control group in published research works, they secure lower false-positive ratios, but they do not form complete representative set of harmless software, since many different software products are available and some of them may even resemble malware in some of their features. This idea initiated our first experiments targeted at packing and related properties of programs [1][2]. We discovered that harmless software shares occurrence of packing with malware, together with other related properties. This led us to deeply examine malicious and harmless samples and search for hidden relations.

Our experiment presented in this paper is unique by means of samples selection focused on specific software domain – system utilities and maintenance tools. By this rather narrow selection we aim to eliminate the influence of usage domain of

This work has been supported by the Slovak Research and Development Agency under the contract No. APVV-15-0055 and by project KEGA no. 079TUKE-4/2017.

software which may play important role in exhibited behaviour and properties. In this way we can better compare malicious and harmless samples and look for features that may help in distinguishing them.

## II. BACKGROUND OF THE EXPERIMENT

Detecting that a program is packed is the first step towards its in-depth analysis [3]. Execution of packed program is often inevitable for recovering original program's code and unveiling its behaviour. Dynamic inspection of malware poses a risk that it may escape from analytic environment and spread on more systems, therefore strict security precautions need to be taken when executing malware.

### A. Packing as Analysis and Detection Prevention

Packers employ compression for reduction of program's size and encryption to obstruct program's reverse-engineering [4]. The resulting file comprises an unpacking routine and packed data blocks. When packed program is executed, the unpacking routine recovers the original program code into memory and directs the execution flow to execute it. Program code can be retrieved by virtual machine monitors or emulators [5] but researchers also look for static techniques to distinguish packed malware from goodware [6][3]. Packers are popular for hindering signature-based detection and static analysis.

A general belief is that packing, together with obfuscation, is a common trait in malicious programs and this idea is repeated among malware analysts and researchers [7][8]. Despite that, it is not easy to find current and accurate rates of packer detections. According to Cisco Blog, they estimate around 70-80% of malware is packed and only around 5% of harmless software is modified by packers<sup>1</sup>. Considering the year the article was published (2010) the rates are now outdated, however, a more recent blog article on Malwarebytes from October 31, 2017 states that "*over the last quarter, we've seen an increase in malware using packers, crypters, and protectors*" and "*the growing number of malware authors using these protective packers has triggered an interest in alternative methods for malware analysis*"[9]. As it seems, packers are not on the decline yet, so investigating their occurrence in harmless software may lead to interesting insights.

<sup>1</sup>[https://blogs.cisco.com/security/malware\\_validation\\_techniques](https://blogs.cisco.com/security/malware_validation_techniques)

### B. Introducing Research Hypotheses

Packing is deemed typical for malicious software but results that would confirm and explain reliability of this assumption are not present. We will try to shed some light on this problem and see if assumptions regarding packing match the reality. Our research is evaluated with statistical tests of null hypothesis and two alternative hypotheses for each of analysed features: amount of detected packers, amount of program's sections, entropy of section *.text*, entropy of section *.rest*, and percentage of program's code in section *.rest*.

The null hypothesis reflects the assumption that values detected – low or high – are not related to malicious or harmless origin of samples, so no significant difference in values will be observed in data sets:

*Hypothesis 0 (H0)*: The difference of values measured for analysed feature in harmless software and malicious software is small and insignificant.

Alternative hypotheses reflect the expectation of higher or lower values in malicious samples:

*Hypothesis 1 (H1)*: Analysed feature has higher values in malicious software when comparing to harmless software.

*Hypothesis 2 (H2)*: Analysed feature has lower values in malicious software when comparing to harmless software.

According to hypothesis H2, packing and related features considered typical for malicious software may be detected in harmless software with higher values. Proving the hypothesis for these features may unveil hidden complications in detection mechanisms based on typical malware features.

For statistical analysis, we used two-sample Wilcoxon rank sum test (U test) for comparing the data sets with confidence level 95% ( $\alpha = 0.05$ ).

## III. RESEARCH DATA AND METHODS

### A. Experimental Sets

The set of experimental samples consisted of utility software distributed on the internet for free. Parsons and Oja explain that utility software is a kind of software purposed to assist with monitoring and configuring a computer system and its software [10]. Utility software covers various maintenance tasks, e.g. deleting temporary files, broken links removal, searching for duplicate files, tasks management, memory optimization or personal files encryption. We targeted this specific group of software because of operations that these programs are designated to perform.

We assume that software which legitimately accesses registry entries, processes, file system, etc., may be a promising target for malware writers which create malicious imitations of the original harmless software. With this in mind we aimed at comparing harmless system maintenance tools and their fake malicious counterparts.

We assembled two experimental sets: One containing legitimate applications in a form of executable files (.exe) downloaded from the internet and another containing 100 samples verified as malicious, collected from malware analytic services. We could not obtain all malware .exe files to

analyse them on our own, therefore to unify our resources we used reports from analysis of samples, which were available for both sets. Also, we could not obtain exact malicious counterparts of all harmless programs, but nevertheless we preserved the domain of malicious samples in utility software. We obtained malware in the domain of utility software by looking at application name, e.g. "win defrag", which hints on the intended purpose of the sample. We also focused on high amount of detections by anti-virus (AV) engines and manually selected samples which met our criteria. Data fields and their extraction are described in Section III-C.

### B. Analytic Tools

We performed our experiments with various kinds of tools:

**PEiD** is a tool that allows to identify packers which have been used on programs' code <sup>2</sup>. Packing and encrypting libraries are often used by malware for concealment of suspicious parts of a program and evasion of detection based on malware signatures. PEiD performs the search based on signature-like definition of several hundreds of known packers. While it is reliable to detect commonly used packers, it may fail on custom-made packers whose signatures are yet unknown. The original web page of the tool is discontinued and according to reports it may have been taken over by malicious actors. We obtained our copy of the tool with REMnux <sup>3</sup> distribution for malware analysis.

**UPX** is a packing tool for executable files <sup>4</sup> which is used in the experiment to check whether tested file is packed or not, and to unpack files that are packed. As Davis *et al.* state [11] in their book, numerous computer viruses use specifically UPX packer. A recent case of its malicious usage is presented in blog article by Nick Biasini *et al.* in which cryptocurrency-mining malware *Dark Test* uses UPX as one of its hiding techniques [12]. The packing problem is discussed also in the work of Guo, Ferrie and Chiueh [7]. Therefore, detection of UPX packer being used on analysed sample arises suspicion.

**VirusTotal** is online malware analysis service. We used it to obtain properties and behaviour of malicious and harmless set of samples <sup>5</sup>. In case of harmless samples it was used for safe and reliable analysis of samples that we collected. In case of malware, since we did not possess original malicious files, we used the service to search for reports from analysis by hash codes of samples that we collected beforehand. Reports generated from analysis contain various information, regarding our experiment e.g. scan results form over 50 anti-virus (AV) solutions, detection of packers by analytic tools F-PROT and PEiD, information from PE header, PE sections with their names and properties, and behavioural information with executed system calls.

Tools for static analysis – PEiD, UPX – run as terminal applications which accept arguments that modify settings and set input and output of analysis. This allowed us to create a

<sup>2</sup>PEiD tool: <https://www.aldeid.com/wiki/PEiD>

<sup>3</sup>REMnux: <https://remnux.org/>

<sup>4</sup>UPX: <http://upx.sourceforge.net/>

<sup>5</sup>VirusTotal: <https://www.virustotal.com>

helper program which utilised these features for automation of analysis.

### C. Experimental Procedure

The experiment was performed in two stages: First we analysed harmless set of programs and evaluated results. In the second stage we proceeded with examining malicious samples. The procedure differed for harmless and malicious samples, mainly because original malicious samples were not available, therefore, the second stage leaned on data provided by analytic reports produced by VirusTotal.

In the first stage we employed our helper program which was developed prior to the experiment for automating the usage of analytic tools. Each harmless sample was analysed by PEiD and UPX to check whether it is packed, and if yes, to identify used packers. Results were collected and summarised in a table. Analysis by VirusTotal followed. We collected produced reports and extracted information of interest.

The second stage regarded malicious samples and employed data extracted from reports of analysis obtained by VirusTotal.

Data obtained from analytic reports comprise the following:

- *Detection results of malware scanners.* In case of positive detection, we obtained name of detection signature, for each scanner separately. We summarised the data as quantity of detections per sample.
- *Detection of packers applied to pack analysed sample.* We acquired names of detected packers and summarised the data as quantity of detected packers per sample.
- *Names of program's sections.* We counted the amount of sections and also stored their names for further manual research. Too few or too many sections comprising the executable file suggest that it was packed, encrypted or otherwise modified in order to disguise original code structure<sup>6</sup>.
- *Entropy of program's sections .text and .rest.* A section usually named as *.text* or *.code* contains program's instructions. In some occasions parts of code may occur out of usual sections, in so-called *.rest*. Presence of this quasi-section is characteristic for programs modified with some packer. Entropy of these sections may show whether they were modified by packing or encryption, which typically cause entropy to be very high. Values are measured in the interval  $< 0,8 >$ . Bytes of program's code have some non-random distribution and therefore low entropy. The higher the value, the more random distribution of bytes, suggesting uncommon modifications.
- *The amount of program's code in section .rest.* We calculated percentage amount of bytes in this section. Large portions of code in this section are typical for packed programs.

To objectively evaluate differences between data of malicious and harmless samples we used statistical analysis,

<sup>6</sup>More information regarding PE file sections: <https://docs.microsoft.com/en-us/windows/desktop/Debug/pe-format>

CI	code injection	HG	HTTP GET request
DLL	runtime DLL	HP	HTTP POST request
DNS	DNS request	MC	mutex created
FC	file created	MO	mutex opened
FD	file deleted	PC	process created
FM	file moved	REG	registry entry
FO	file opened	SS	service started
FR	file read	SW	searched window
FW	file written	TCP	TCP data flow
HOOK	hooking activity	UDP	UDP data flow

TABLE I  
BEHAVIOURAL CATEGORIES AND THEIR ABBREVIATIONS.

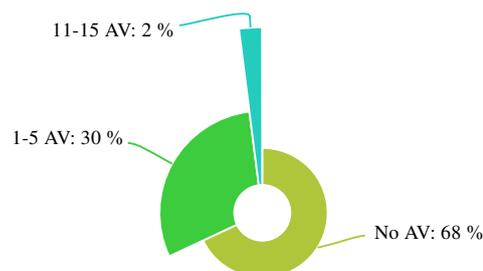


Fig. 1. Pie chart with amount of *harmless samples* that were positively detected by given amount of AV scanners.

specifically two-sample *Wilcoxon rank sum test (U-test)* with confidence level 95% ( $\alpha = 0.05$ ).

From reports provided by VirusTotal we were able to obtain high-level information about behaviour of analysed samples as a list of executed system calls. We used the data in summative way as quantities of operations in behavioural categories listed in Table I. Each sample was then described by 20 numerical values representing occurrences of 20 types of behaviours.

## IV. RESULTS AND OBSERVATIONS OF THE EXPERIMENT

The following sections present results regarding analysed features and statistical analysis (end of this section, Table II).

### A. Detection Results of Malware Scanners

During our experiment VirusTotal employed usually 56 AV scanners but in some cases, for unknown reasons, few of them were unavailable in the report.

Pie charts in Fig. 1 and 2 show amounts of AV scanners (height of a slice) that positively detected analysed samples (their amount as width of a slice), grouped into ranges for improved visual clarity. Among harmless samples no detection (Fig. 1) prevails, but some were detected as threat nevertheless. Among malware samples (Fig. 2) only one had no detection and the rest of them was detected by multitude of AV scanners.

### B. Detection of Packer Usage

Pie charts in Fig. 3 and 4 show the amounts of detected packers. Height of a slice represents the amount of packers detected in single sample and width shows the amount of samples detected to be packed with given amount of packers.

Only 20% of *harmless samples* were detected as not packed, the rest was modified by packers ranging from 1 to 7.

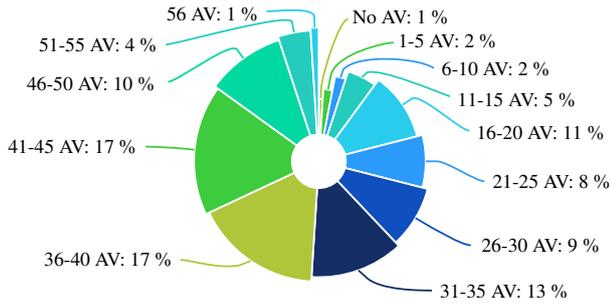


Fig. 2. Pie chart with amount of *malicious samples* that were positively detected by given amount of AV scanners.

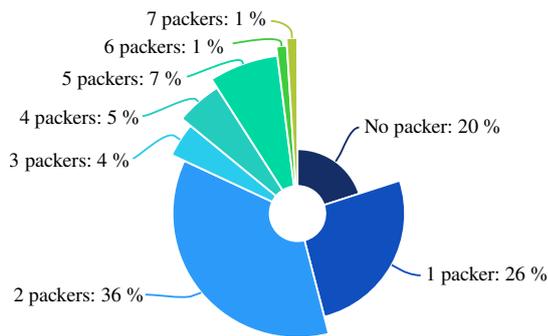


Fig. 3. Pie chart with amount of detected packers in *harmless samples*.

Data from malware samples present more surprises. One malware sample was packed 46 times – that is extreme. However, majority of samples was detected as not packed at all – another surprise, since usual expectations are that malware will be packed massively, not the opposite.

Regarding distribution of amount of packers detected, we can see in Fig. 5 that for *malware* the value of median matches the lower (first) quartile (value 0), and in case of *harmless software* median matches the upper (third) quartile (value 2). From this we can deduce that harmless samples are more prone to being packed, at least with packers that are detectable by available tools. Outliers (extreme values) are not shown in order to prevent plot deformation.

Regarding the hypothesis H1 saying that values of analysed feature – occurrence of packers – is higher in malicious software than in harmless software (Table II, row *Packers amount*, alternative *Higher*), the U-test resulted with p-value  $> 0.99$  which by far exceeds the significance level. As a result, we fail to reject the null hypothesis for this case.

For alternative hypothesis H2 suggesting that occurrence of packers is lower in malicious software, thus prevails in harmless software (Table II, row *Packers amount*, alternative *Lower*), the U-test resulted with p-value  $7.3124 \times 10^{-12}$  which is far below the significance level. As a result, we reject the null hypothesis and accept the alternative hypothesis H2.

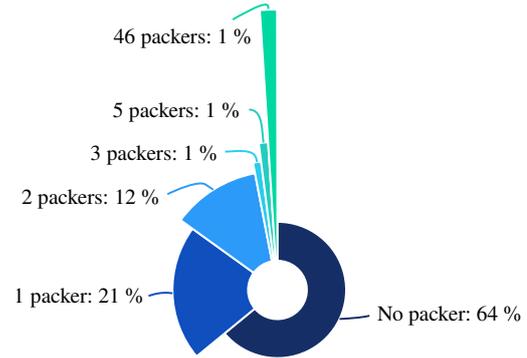


Fig. 4. Pie chart with amount of detected packers in *malicious samples*.

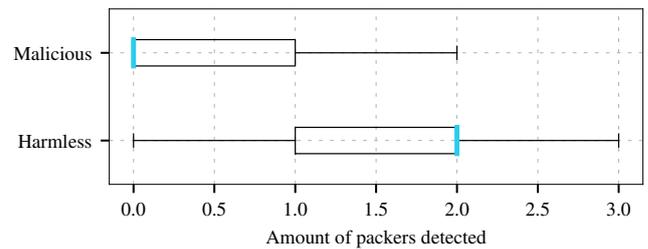


Fig. 5. Boxplots for amounts of packers detected in *harmless* and *malicious* samples. Outliers are not shown in the figure.

### C. Amount of Program's Sections

Pie charts in Fig. 6 and 7 show the amounts of detected sections in analysed programs. Height of a slice represents the amount of sections detected in single sample and width shows the amount of samples with given amount of sections. The amount of sections in majority of harmless samples is quite high – 8 – but the amount in malware is lower.

For *harmless samples*, the interquartile range is higher (Fig. 8) – the box is much wider, in comparison to *malware samples*. Both sets of samples match on the first quartile with value 4. There are 6 *harmless samples* with no section detected. This may be caused by different actual file format than PE so the section table could not be retrieved. While several outliers among *malware samples* may make the impression that the amount of sections is high in malware, values of median clearly show that *harmless samples* tend to have higher amount of sections.

For alternative hypothesis H1 saying that amount of sections is higher in malware than in goodware (Table II, row *Sections amount*, alternative *Higher*), the U-test resulted with p-value  $> 0.99$  which by far exceeds the significance level. As a result, we fail to reject the null hypothesis for this case.

For alternative hypothesis H2 saying that amount of sections is lower in malware, so prevails in harmless software (Table II, row *Sections amount*, alternative *Lower*), the U-test resulted with p-value  $1.9375 \times 10^{-6}$  which is far below the significance level. As a result, we reject the null hypothesis and accept the alternative hypothesis H2 – sections prevailing in goodware.

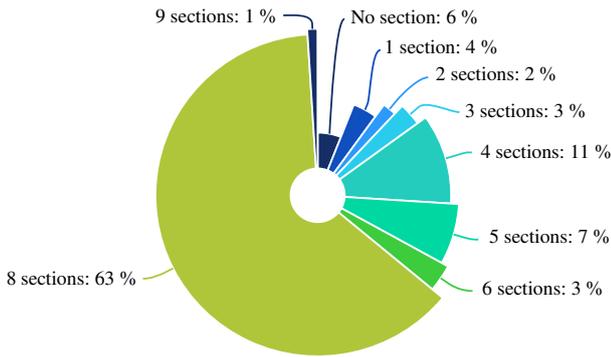


Fig. 6. Pie chart with the amount of detected sections in *harmless samples*.

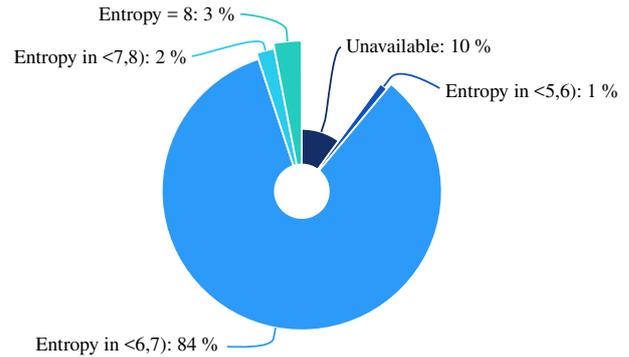


Fig. 9. Pie chart with entropy of section *.text* in *harmless samples*.

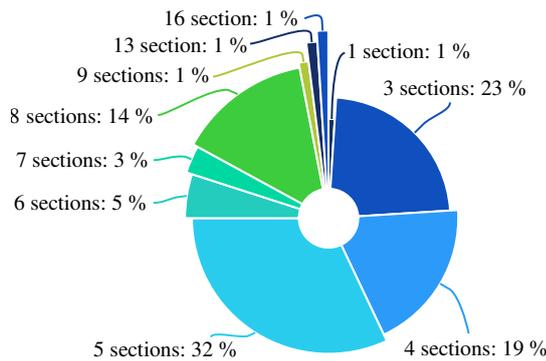


Fig. 7. Pie chart with the amount of detected sections in *malware samples*.

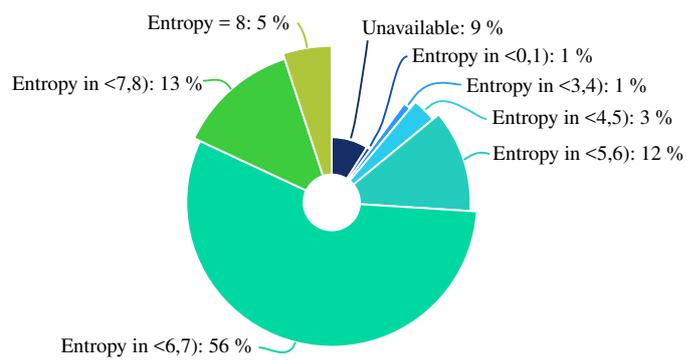


Fig. 10. Pie chart with entropy of section *.text* in *malware samples*.

*D. Entropy of section .text*

Section *.text* (or *.code*) contains executable instructions, therefore its entropy should not be normally very high.

Pie charts in Fig. 9 and 10 show measured entropies. Height of a slice represents range of values of entropy and width shows the amount of samples with given entropy value. We can see that both malicious and harmless group have majority of samples with entropy in range < 6,7). In *harmless samples*, other ranges of entropy occur seldom – only in 6 samples – and the rest comprises samples in which the section could not be precisely identified. *Malware* shows wider variety of entropy, both on the lower and higher spectrum of value ranges.

Boxplot (Fig. 11) shows that inter-quartile range is wider in *malware* and extremes are much more apart. Medians,

however, are close to 6.5 in both malicious and harmless samples. This suggests that the difference in values may not be significant regarding *.text* section entropy.

For alternative hypothesis H1 that values of *.text* section entropy are higher in malware than in goodware (Table II, row *.text entropy*, alternative *Higher*), the U-test resulted with p-value 0.6556 which by far exceeds the significance level. We fail to reject the null hypothesis for this case.

For alternative hypothesis H2 that values of *.text* section entropy are lower in malware (Table II, row *.text entropy*, alternative *Lower*), the U-test resulted with p-value 0.3453 which also exceeds the significance level. We fail to reject the null hypothesis and conclude that differences in values of entropy of section *.text* are not statistically significant.

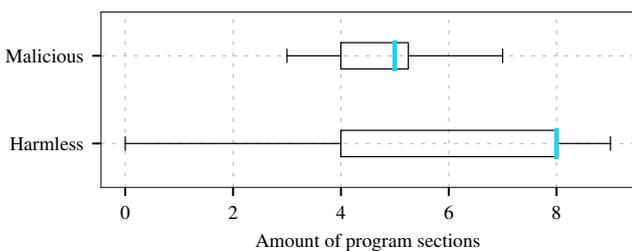


Fig. 8. Boxplots for amount of program's sections in *harmless* and *malicious* samples. Outliers are not shown in the figure.

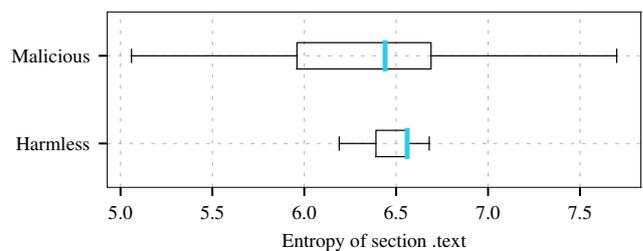


Fig. 11. Boxplots for entropy of section *.text* in *harmless* and *malicious* samples. Outliers are not shown in the figure.

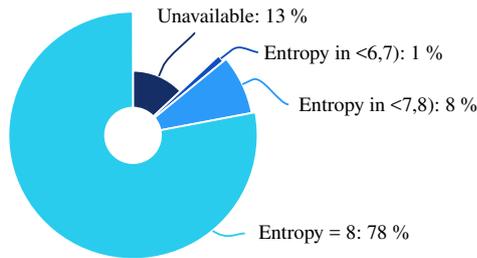


Fig. 12. Pie chart with entropy of section *.rest* in harmless samples.

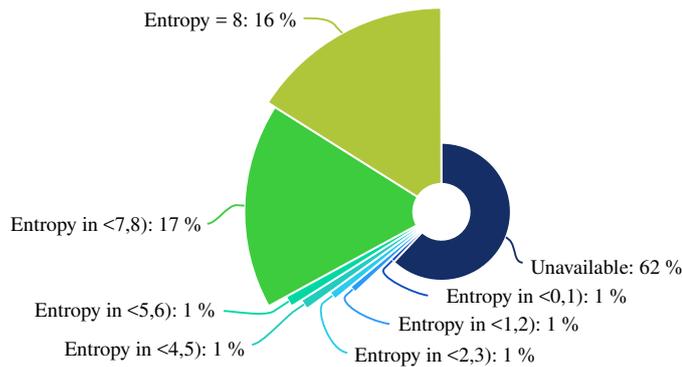


Fig. 13. Pie chart with entropy of section *.rest* in malware samples.

### E. Entropy of section *.rest*

Since *.rest* is not a program section *per se*, it may not be detected in many programs. On the other hand, its presence suggests that special measures have been made to conceal (at least from plain sight) some incriminating portions of code. It comes naturally to assume presence of *.rest* in majority of malware samples, but looking at our previous experiments, assumptions about malware can be quite misleading.

Pie charts in Fig. 12 and 13 show measured entropy of section *.rest*. Height of a slice represents range of values of entropy and width shows the amount of samples with given entropy value. In case of *harmless samples*, 13 programs had no detectable code out of common sections but the rest of them had, with measured entropy from 6 to 8 – which is a maximum. Contrary to that, *malware samples* showed no section *.rest* in majority of cases (62 samples), and in case of its presence, entropy reached value from 6 to 8 only in 33 samples.

Lower quartile, upper quartile and median for *harmless samples* meet at value 8 (Fig. 14) and contrast with values of malware – it has median at zero, mostly due to absence of the section.

For alternative hypothesis H1 that values of *.rest* section entropy are higher in malware than in harmless software (Table II, row *.rest entropy*, alternative *Higher*), the U-test resulted with p-value  $> 0.99$  which by far exceeds the significance level, so we fail to reject the null hypothesis.

For alternative hypothesis H2 that values of *.rest* section entropy are lower in malware (Table II, row *.rest entropy*, alternative *Lower*), the U-test resulted with p-value  $2.7074 \times 10^{-18}$  which is far below the significance level. We reject the null

hypothesis and accept the alternative hypothesis H2 that *.rest* section entropy is lower in malware than in goodware.

### F. Percentage of code in section *.rest*

Percentages that were found among analysed *harmless samples* are shown in Fig. 15 in ascending order.

Surprisingly, large programs' size present in section *.rest* prevails in *harmless software*. We suppose that it may be inflicted by some commonly used packers and application building tools, such as INNO. However, further research needs to be made to confirm this opinion.

Examination of *malware* showed that *.rest* manifested in much fewer samples than in the harmless set. This result corresponds with findings from analysis of entropy of section *.rest* (Sec. IV-E). The percentage of *.rest* section in file's size (Fig. 16) was calculated with data obtained from VirusTotal analytic reports.

62 *malware* samples contained no data outside sections listed in PE file header and so section *.rest* was confirmed to be absent in them. The case of 90% or more of program's bytes was present only in 11 samples. Again, this strongly contrasts with results from harmless samples.

Figure 17 shows that for *harmless samples* the first and the third quartile have high values and samples with less than 80 % percent of code in *.rest* are basically outliers. Quartiles of malware data contrast to that as low – the first quartile and median match at value 0. The boxplot suggests notable differences in values of malicious and harmless samples.

For alternative hypothesis H1 that percentages of code in *.rest* section are higher in malware than in goodware (Table II, row *.rest percentage*, alternative *Higher*), the U-test resulted with p-value  $> 0.99$  which by far exceeds the significance level. As a result, we fail to reject the null hypothesis.

For alternative hypothesis H2 that percentages of code in *.rest* section are lower in malware (Table II, row *.rest percentage*, alternative *Lower*), the U-test resulted with p-value  $3.2413 \times 10^{-17}$  which is far below the significance level. As a result, we reject the null hypothesis and accept the alternative hypothesis H2 about percentage of code in section *.rest* being lower in malware than in harmless software.

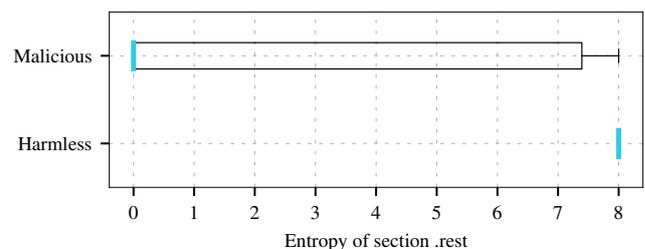


Fig. 14. Boxplots for entropy of section *.rest* in harmless and malicious samples. Outliers are not shown in the figure.

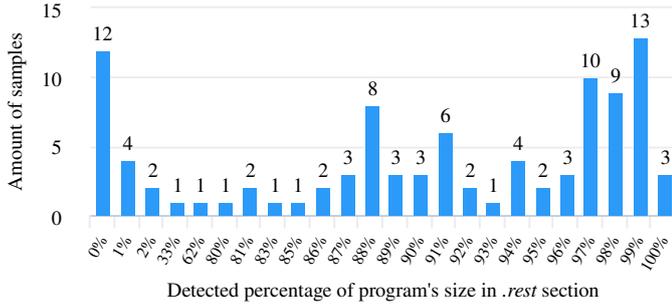


Fig. 15. Measured percentages of .rest section from total size of code in harmless samples, and number of samples with that percentage of code in .rest section.

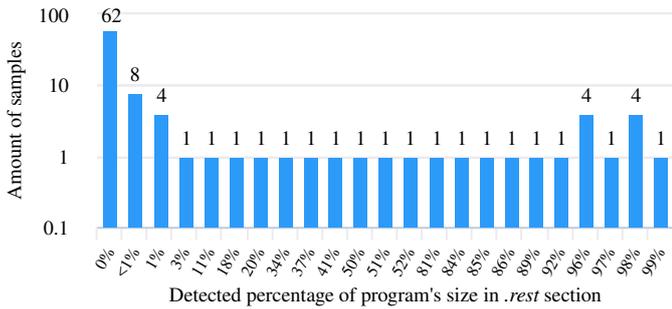


Fig. 16. Measured percentages of .rest section from total size of code in malicious samples, and number of samples with that percentage of code in .rest section. The y-axis with amounts of samples is scaled logarithmically due to great differences among values.

G. Relationships Between Analysed Features

We explored also relationships between analysed features by calculating Pearson's correlation coefficients for each pair of features that were discussed in statistical analysis in previous section. The measure of correlation is described by values from interval  $(-1, 1)$ , with the following interpretation:

- value  $-1$  represents perfect negative correlation, i.e. when values of one feature are high, second feature's values are low, and vice versa,
- value  $0$  represents no measurable linear correlation,
- value  $1$  represents perfect positive correlation, i.e. when value of one feature is high, so is value from the second feature, and vice versa.

Visual representation of correlation matrix as a "heatmap" (Fig. 18, 19) can aid in understanding notable linear relationships between features. We can see that in heatmaps of both malicious and harmless samples only positive correlations were present. For harmless samples the most notable correlation is between amount of detected sections and entropy of section .rest. It seems that when packers are used for hiding program's code the amount of sections is decreased and large portions of code are then in section .rest. We explored the issue further and discovered that it occurred mainly when a sample was packed by packer UPX. If it was the only packer used, it placed all the code into section named .upx.

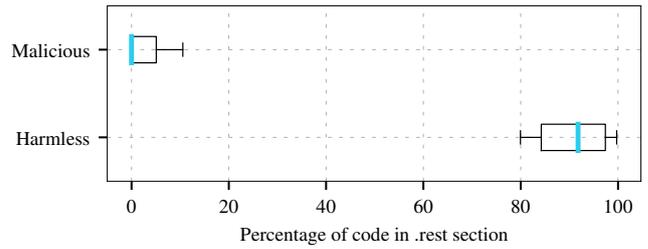


Fig. 17. Boxplots for percentage of program's code in section .rest in harmless samples and malicious samples. Outliers are not shown in the figure.

Data sets		Alternative	Result
Malicious	Harmless		
Packers amount	Lower		$7.3124 \times 10^{-12}$
Packers amount	Higher		$\approx 1$
Sections amount	Lower		$1.9375 \times 10^{-6}$
Sections amount	Higher		$\approx 1$
.text entropy	Lower		0.3453
.text entropy	Higher		0.6556
.rest entropy	Lower		$2.7074 \times 10^{-18}$
.rest entropy	Higher		$\approx 1$
.rest percentage	Lower		$3.2413 \times 10^{-17}$
.rest percentage	Higher		$\approx 1$

TABLE II  
SUMMARY OF RESULTS FROM STATISTICAL ANALYSIS MADE WITH Wilcoxon rank sum test (U-test) WITH CONFIDENCE LEVEL 95% ( $\alpha = 0.05$ ). THE ALTERNATIVE DESCRIBES RELATION BETWEEN VALUES OF MALICIOUS AND HARMLESS SET OF SAMPLES.

Another notable correlations in harmless samples were between amounts of sections and detected packers, and between percentage of code in section .rest and entropy of this section, but values for these correlations are not that high.

Regarding malicious samples, only one correlation is notable and it was measured between values of percentage of code in section .rest and entropy of this section. This relates to finding that numerous malware samples do not have this section so presence of this section with some entropy will cause this correlation to be high.

H. High-Level Behaviours

Beside features related to packing we recorded and measured also quantity of executed system calls which belong to behavioural categories listed in Table I. After observing the values we realised that much more possibilities of analysis are opening ahead of us and thus will require further work. However, to supplement findings from previous sections with at least several interesting insights about behaviour, we analysed correlations between pairs of behaviours and created heatmaps for malicious (Fig. 20) and harmless samples (Fig. 21).

First notable thing is that harmless samples had no occurrence in behavioural categories CI - code injection and FC - file created, therefore lines for these features are blank in Fig. 21.

We can see that heatmap of harmless samples contains more high correlations than heatmap of malware samples. Some of

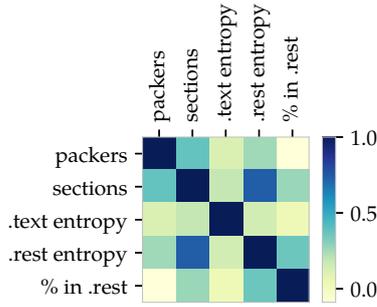


Fig. 18. Correlation matrix heatmap for properties related to packing in *harmless samples*.

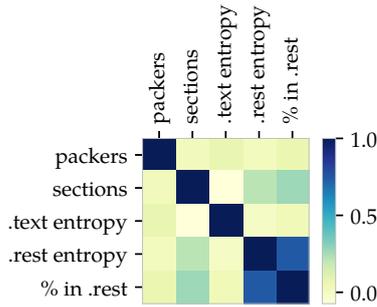


Fig. 19. Correlation matrix heatmap for properties related to packing in *malicious samples*.

them match but for *harmless samples*, these high correlations between pairs of features are unique:

- *FM* - file moved and *FD* - file deleted,
- *FO* - file opened and *FM* - file moved,
- *FO* - file opened and *FD* - file deleted,
- *FW* - file written and *FM* - file moved,
- *FW* - file written and *FD* - file deleted,
- *HG* - HTTP GET request and *DNS* - DNS request,
- *MO* - mutex opened and *MC* - mutex created,
- *SW* - searched window and *MC* - mutex created,
- *TCP* - TCP data flow and *DNS* - DNS request.

It seems that for *harmless samples* operations related to files are often performed and quantitatively relate to each other.

For *malicious samples*, interesting correlations are between following features:

- *MC* - mutex created and *FD* - file deleted,
- *MO* - mutex opened and *DLL* - runtime DLL,
- *REG* - registry entry and *FO* - file opened,
- *REG* - registry entry and *FR* - file read,
- *SS* - service started and *FD* - file deleted,
- *UDP* - UDP data flow and *HG* - HTTP GET request.

*Malware samples* seem to be focused more on mutex, registry and service operations and they are performed similarly often as various operations with files, mainly opening, reading and deleting.

### I. Summary

Entropy of section *.text* was notably high in many malicious and harmless samples, so it seems that concealment is targeted

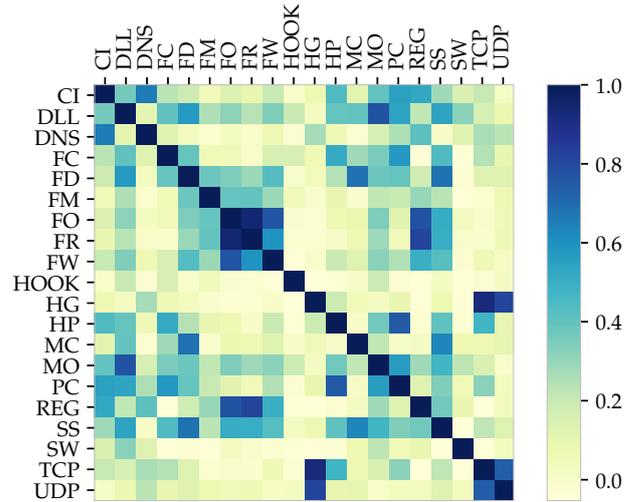


Fig. 20. Correlation matrix heatmap for high-level behaviours of *malicious samples*.

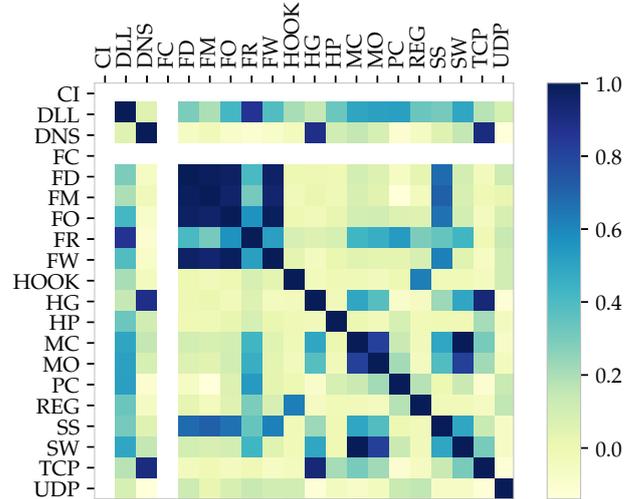


Fig. 21. Correlation matrix heatmap for high-level behaviours of *harmless samples*.

at program's section with executable instructions. Difference between sets resulted as insignificant, but lower amount of detected packers in malware suggests that custom packers are used, which are undetectable by common analytic tools. Section *.rest* was not detected in majority of malware samples. This suggests that malware writers use custom code concealing techniques that do not create this quasi-section.

Regarding the assumptions about malware being packed, they are probably true, however, common tools seem now insufficient for malware packers detection. Information pointing at programs being packed need to be collected from deeper levels of analysis: Amounts and names of packers provided by analytic tools may be incomplete, but dissecting a program into sections and observing their entropy shows the missing

pieces of information.

Correlations between high-level behaviours hint some interesting relationships between types of operations performed during execution of programs and it would be interesting to analyse specific cases of behaviours that are executed together. Also, it would be beneficial to look at behaviours of malware and harmless software from another software domain and examine if there are some repeating patterns present in values.

Insights obtained from experiments that we performed are significant regarding research and implementation of novel detection systems based on machine learning and neural networks, since they utilise quantitatively measurable features of software samples. Our results present options for feature engineering to obtain high-quality training data: Neglecting features that are insignificant or combining closely related features to one for reducing dimensionality of data. Mindful selection of training data will allow to direct more attention and time to selection and experimentation with detection algorithms.

Based on our work new research questions arise. Behavioural and non-behavioural properties of programs vary greatly and as we "zoom in" to specific software domain of malicious and harmless samples, we see that usual assumptions cease to be applicable. For example, results of signature-based detection of packers do not agree with expectations towards malware and goodware (Sec. IV-B) and are therefore unsuitable for machine learning purposes. Clearly, malware writers made it harder to detect packers, so deeper analysis of samples is required to obtain desired data. This opens up a question whether now-so-popular utilisation of massive data is not in fact a limitation in improving malware detection, and we ought to pursue more specific and narrow features for differentiating malware from goodware instead.

## V. DISCUSSION

Beside reports from analysis obtained from VirusTotal we used tools of static analysis in this experiment for the following reasons:

- 1) Although dynamic analysis compared with static analysis indicates smaller issues with malware obfuscation, static analysis allows to detect malicious features which seem to occur randomly during a program execution or in a specific execution environment.
- 2) We could automate the analytic process for large amount of experimental samples in our custom created utilities.
- 3) Automated process of static analysis is less time consuming compared with dynamic analysis, which requires execution of each analysed sample.

Concerning point 1, every program can comprise numerous execution paths, also called *execution traces*. The disadvantage of dynamic analysis is that only one execution trace can be observed at a time. Concerning several traces, also static analysis with reverse engineering is problematic. However, Beaucamps *et al.* addressed this problem in their work [13] and proposed a method for static analysis of execution traces

acquired from control-flow graphs. Macedo and Touili also discuss the issue in their work [14].

UPX packer is commonly known as packing tool often misused for covering malicious code. Marak states, however, that obfuscating effects of the packer are not among its original features and result from altering its original code [15]. A look at licence of UPX packer reveals that modifications and usage of the packer in such way is violating the rules of tool's legal usage <sup>7</sup>. In fact, UPX packer should not be suspicious by itself, like many blogs and papers state, but the illegal modifications made to it are what causes trouble. This fact should be given more attention and researchers should avoid improper simplifications of the matter.

In some cases, PE sections names may reveal name of a packer used for concealing program's code, however, this information is not fully reliable since section names can be modified by various tools, e.g. by PE Explorer <sup>8</sup>. In our observations we also encountered section names being some gibberish or blank—obviously someone removed the original section name on purpose.

### A. Related Work

Malware signatures have still very important role in malware detection, although their effectiveness on malicious samples concealed by techniques that alter syntactic form of a program is questionable. What is more, with rapidly growing number of new malware samples the extraction of signatures requires a lot of precious time. Griffin *et al.* addressed this problem and presented a system for automated generation of malware signatures [16]. An interesting part of their work describes features which they analysed in malicious programs. Concerning syntactic form of a program authors mention patterns emerging in operational code which may represent precursors of non-standard or suspicious behaviour of the program:

- Constant values like IP addresses, email addresses,
- access to memory with unusual offset,
- local function calls, non-library function calls, context of a function call and used parameters,
- suspicious mathematical operations which may indicate obfuscation.

These patterns are used for refining potential malware signatures through, as they call it, *code interestingness check*. In our research they served as an inspiration for comparing features of malicious and harmless programs and looking for patterns which could be employed as indicators of malicious intentions.

### B. Influences on the Study and its Outcomes

Results of our research showed that for special-purpose software packing may be detected more often in harmless samples than in malicious samples, which is in contrast with common assumptions about malware. Nevertheless, several factors could have affected the outcomes even when we made an effort to mitigate them as best we could.

<sup>7</sup>UPX licence: <https://upx.github.io/upx-license.html>

<sup>8</sup>PEexplorer: [http://www.heaventools.com/PE\\_Explorer\\_section\\_editor.htm](http://www.heaventools.com/PE_Explorer_section_editor.htm)

- *Selection of samples.* Commercial paid software was not included in the study but its properties may have been different from what we found in harmless samples. However, obtaining numerous samples of paid maintenance software was not feasible in our research project. Concerning malware samples, it is hard to trace their origin since we worked just with reports from their analysis, not with samples directly. This may have also considerable effect.
- *Collection of samples.* Samples were found on the internet by search engine with specific keywords. Different keywords may have led to different outputs, even when we tried to explore as many various results as we could.
- *The usage domain.* Samples that we experimented with belong to system utility software. Samples from different domains may have different properties regarding packing.
- *Analytic tools.* The tools that we used in our study to gather data of interest do not guarantee 100% correctness of data. There is a chance that detecting fewer packers among malicious samples is caused by inability of tools to unveil usage of hidden, sophisticated custom packer developed by malware creators. This problem, however, is not in our power to mitigate.
- *Other errors.* Several samples had no program sections detected. This may have been caused by an unknown error during analysis performed by tools we used or by difference of actual file format from the format declared by the sample.

## VI. CONCLUSION

We presented a different, novel approach to malware research that is based on narrow selection of experimental samples from specific software domain and statistical evaluation of differences between malicious and harmless software. Several ideas inspired us to perform this experiment:

- 1) Packing is often applied to malicious software with intent to obstruct reverse-engineering, hinder static analysis, and hide incriminating code from malware detectors.
- 2) Although packing is typical for malware, it may be used also on harmless software for protection against intellectual property theft.
- 3) In research circles, a discussion about distinguishing malicious packing from harmless one regards syntactic features of program's operational code, e.g. bytes distribution, entropy, data in so-called *.rest* section.
- 4) Harmless programs have not been given appropriate attention, especially from the context of features relevant for distinguishing between malicious and harmless case of packing, and their reliability.

Although packers are massively used by malware creators, they are also applied for protection of intellectual property in harmless software, making it complicated to separate bad and good intentions behind packer's usage.

In the paper we showed that differences in values between malicious and harmless programs are significant regarding amount of detected packers, amount of program sections,

percentage of code in section *.rest* and its entropy. Entropy of section *.text* together with amounts of packers detected suggest that malware writers create custom packers that are nearly undetectable by common analytic tools.

It is necessary to keep in mind that results presented here concern samples from the domain of maintenance and utility software and samples from other domains may yield different results. In that case, however, it would be interesting to research the influence of software domain selection on values of analysed features, since it may be significant.

## REFERENCES

- [1] J. Št'astná and M. Tomášek, "Exploring malware behaviour for improvement of malware signatures," in *IEEE 13th International Scientific Conference on Informatics, 2015*, Nov 2015. doi: 10.1109/Informatics.2015.7377846 pp. 275–280.
- [2] J. Št'astná and M. Tomášek, "The problem of malware packing and its occurrence in harmless software," *Acta Electrotechnica et Informatica*, vol. 16, no. 3, pp. 41–47, 2016. doi: 0.15546/aeeci-2016-0022
- [3] T.-Y. Wang and C.-H. Wu, "Detection of packed executables using support vector machines," in *International Conference on Machine Learning and Cybernetics (ICMLC), 2011*, vol. 2, 2011. doi: 10.1109/ICMLC.2011.6016774. ISSN 2160-133X pp. 717–722.
- [4] S. Josse, "Secure and advanced unpacking using computer emulation," *Journal in Computer Virology*, vol. 3, no. 3, pp. 221–236, 2007. doi: 10.1007/s11416-007-0046-0
- [5] M. Šipoš and S. Šimoňák, "Rasp abstract machine emulator – extending the emustudio platform," *Acta Electrotechnica et Informatica*, vol. 17, no. 3, pp. 33–41, 2017. doi: 0.15546/aeeci-2017-0024
- [6] G. Jacob, P. Comporetti, M. Neuschwandtner, C. Kruegel, and G. Vigna, "A static, packer-agnostic filter to detect similar malware samples," in *Detection of Intrusions and Malware, and Vulnerability Assessment*, ser. LNCS, vol. 7591. Springer Berlin Heidelberg, 2013. doi: 10.1007/978-3-642-37300-8\_6. ISBN 978-3-642-37299-5 pp. 102–122.
- [7] F. Guo, P. Ferrie, and T.-c. Chiueh, "A study of the packer problem and its solutions," in *Recent Advances in Intrusion Detection*, ser. LNCS, vol. 5230. Springer Berlin Heidelberg, 2008. doi: 10.1007/978-3-540-87403-4\_6. ISBN 978-3-540-87402-7 pp. 98–115.
- [8] A. Singh and A. Lakhota, "Game-theoretic design of an information exchange model for detecting packed malware," in *6th International Conference on Malicious and Unwanted Software (MALWARE), 2011*, 2011. doi: 10.1109/MALWARE.2011.6112319 pp. 1–7.
- [9] P. Arntz. Analyzing malware by api calls. [Online]. Available: <https://blog.malwarebytes.com/threat-analysis/2017/10/analyzing-malware-by-api-calls/>
- [10] J. Parsons and D. Oja, *New Perspectives on Computer Concepts 2013: Comprehensive*, ser. New Perspectives. Cengage Learning, 2012. ISBN 9781133190561
- [11] M. Davis, S. Bodmer, and A. LeMasters, *Hacking exposed malware and rootkits*. New York: Mc-Graw Hill, 2010. ISBN 978-0-07-159119-5
- [12] N. Biasini, E. Brumaghin, W. Mercer, and J. Reynolds. Ransom where? malicious cryptocurrency miners takeover, generating millions. [Online]. Available: <http://blog.talosintelligence.com/2018/01/malicious-xmr-mining.html>
- [13] P. Beaucamps, I. Gnaedig, and J.-Y. Marion, "Abstraction-based malware analysis using rewriting and model checking," in *Computer Security - ESORICS 2012*, ser. LNCS, vol. 7459. Springer Berlin Heidelberg, 2012. doi: 10.1007/978-3-642-33167-1\_46. ISBN 978-3-642-33166-4 pp. 806–823.
- [14] H. Macedo and T. Touili, "Mining malware specifications through static reachability analysis," in *Computer Security - ESORICS 2013*, ser. LNCS, vol. 8134. Springer Berlin Heidelberg, 2013. doi: 10.1007/978-3-642-40203-6\_29. ISBN 978-3-642-40202-9 pp. 517–535.
- [15] V. Marak, *Windows Malware Analysis Essentials*. Packt Publishing, 2015. ISBN 9781785287633
- [16] K. Griffin, S. Schneider, X. Hu, and T.-c. Chiueh, "Automatic generation of string signatures for malware detection," in *Recent Advances in Intrusion Detection*, ser. LNCS, vol. 5758. Springer Berlin Heidelberg, 2009. doi: 10.1007/978-3-642-04342-0\_6. ISBN 978-3-642-04341-3 pp. 101–120.

# 4<sup>th</sup> International Workshop on Language Technologies and Applications

**D**EVELOPMENT of new technologies and various intelligent systems creates new possibilities for information processing. Natural Language Processing (NLP) addresses problems of automated understanding, processing, evaluation and generation of natural human languages. LTA workshop provides a venue for discussion and presenting innovative research in NLP domain, but not restricted, to: computational and mathematical modeling, analysis and processing of any forms (spoken, handwritten or text) of human language, interactions via Virtual Reality and Augmented Reality, Computational Intelligence models and applications but also other various applications in decision support systems. We welcome papers covering innovative applications and practical usage of theoretical aspects. The LTA workshop will provide an opportunity for researchers and professionals to discuss present and future challenges as well as potential collaboration for future progress in the field.

## TOPICS

The submitted papers shall cover research and developments in all NLP aspects, such as (however this list is not exhaustive):

- Computational Intelligence methods applied to language & text processing
- text analysis
- language networks
- text classification
- language networks, resources and corpora
- document clustering
- various forms of text recognition
- machine translation
- intelligent text-to-speech (TTS) and speech-to-text (STT) methods
- authorship identification and verification
- author profiling
- plagiarism detection
- sentiment analysis
- NLP applications in education
- knowledge extraction and retrieval from text and natural language structures
- multi-modal and natural language interfaces
- innovative language-oriented applications and tools
- interactions models and applications via Virtual Reality and Augmented Reality
- NLP for text analysis in forensic linguistics and cybersecurity

## EVENT CHAIRS

- **Damasevicius, Robertas**, Kaunas University of Technology, Lithuania
- **Martinčić – Ipšić, Sanda**, University of Rijeka, Croatia
- **Napoli, Christian**, Department of Mathematics and Informatics, University of Catania, Italy
- **Sanada, Haruko**, Ritssho University, Japan
- **Woźniak, Marcin**, Institute of Mathematics, Silesian University of Technology, Poland

## PROGRAM COMMITTEE

- **Artiemjew, Piotr**, University of Warmia and Mazury, Poland
- **Bajović, Dragana**, University of Novi Sad, Serbia
- **Burdescu, Dumitru Dan**, University of Craiova, Romania
- **Čukić, Bojan**, UNC Charlotte, United States
- **Dobrišek, Simon**, University of Ljubljana, Slovenia
- **Gelbukh, Alexander**, Instituto Politécnico Nacional, Mexico
- **Harbusch, Karin**, Universität Koblenz-Landau, Germany
- **Ivanović, Dragan**, University of Novi Sad, Serbia
- **Kapočiūtė-Dzikiėnė, Jurgita**, Vytautas Magnus University, Lithuania
- **Krilavičius, Tomas**, Vytautas Magnus University, Lithuania
- **Kurasova, Olga**, Vilnius University, Institute of Mathematics and Informatics, Lithuania
- **Lopata, Audrius**, Vilnius University, Lithuania
- **Madjarov, Gjorgji**, Ss. Cyril and Methodius University in Skopje, Faculty of Computer Science and Engineering, Macedonia
- **Marszałek, Zbigniew**, Silesian University of Technology, Poland
- **Maskeliūnas, Rytis**, Kaunas University of Technology, Lithuania
- **Matson, Eric T.**, Purdue University, United States
- **Meštrović, Ana**, University of Rijeka, Croatia
- **Mikelić-Preradović, Nives**, University of Zagreb, Croatia
- **Nowicki, Robert**, Czestochowa University of Technology, Poland
- **Poław, Dawid**, Institute of Mathematics, Silesian University of Technology, Poland
- **Stanković, Ranka**, University of Belgrade, Serbia
- **Starzewski, Janusz**, Czestochowa University of Technology, Poland

- **Szymański, Julian**, Gdansk University of Technology, Poland
- **Tahmasebi, Nina**, University of Gothenburg, Sweden
- **Tambouratzis, George**, Institute for Language and Speech Processing, Athena Research Centre, Greece
- **Trivodaliev, Kire**, Ss. Cyril and Methodius University in Skopje, Faculty of Computer Science and Engineering, Macedonia
- **Wang, Lipo**, Nanyang Technological University, Singapore
- **Wei, Wei**, School of Computer Science and EngineeringXi'an University of Technology, China

# Multilingual Knowledge Base Completion by Cross-lingual Semantic Relation Inference

Nadia Bebeshina-Clairret  
LIRMM, University of Montpellier,  
860 rue de St Priest 34095 Montpellier  
France  
Email: clairret@lirmm.fr

Mathieu Lafourcade  
LIRMM, University of Montpellier,  
860 rue de St Priest 34095 Montpellier  
France  
Email: lafourcade@lirmm.fr

**Résumé**—In the present paper, we propose a simple endogenous method for enhancing a multilingual knowledge base through the cross-lingual semantic relation inference. It can be run on multilingual resources prior to semantic representation learning. Multilingual knowledge bases may integrate preexisting structured resources available for resource-rich languages. We aim at performing cross-lingual inference on them to improve the low resource language by creating semantic relationships.

## I. INTRODUCTION

HIGHLY structured knowledge bases (KBs) such as lexical semantic networks (LSNs) contain various connectivity patterns that can be learned as node features using dedicated frameworks i.e. node2vec [10]. However, semantic relations are often unequally distributed over such knowledge resources. Some of the language partitions may benefit from integrating structured resources which are more easily available for "rich" languages i.e. Princeton WordNet (PWN) [7], ConceptNet [21], YAGO [22] for English, RezoJDM [12] for French.

Unlike large factual KBs, the LSNs explicitly represent taxonomic relations (*hyponymy*, *meronymy*), predicate-argument relations, typical characteristic, and possibly other relation types (*entailment*, causal relations) as well as polysemy (through synsets, refinements). A meta-information related to the relation weight (power of association i.e. in RezoJDM), a confidence score linked to the origin of the relations integrated from some existing resources (i.e. in ConceptNet), annotation as well as negatively weighted relations that explicitly model "noise" (relation considered as false, i.e. RezoJDM) may be attached to the LSN relations in the framework of a particular model. Thus, automated semi-structured approaches to the multilingual LSN building represent a hard task : when available, models may vary from one language to another. For instance, the modeling of meronymy relations may reflect different vision of this relation type. In ConceptNet, the meronymy is represented as a *hasPart* relation. PWN introduces the distinction between part (*mammal*→*mouth*), substance (*wine*→*alcohol*), and member (*bee*→*bee colony*) meronymy. RezoJDM model includes all the relations covered by PWN and adds the holonymy relation (*cutlet*→*beef*).

## II. STATE OF THE ART

Cross-lingual relationship inference benefits from active research efforts. State of the art inference in KBs include rule-based and machine learning approaches. In the framework of the large KBs such as NELL [3], several approaches centered on the equivalence between entities and relationships have been introduced. For instance, authors in [11] describe the experience of merging several monolingual editions of NELL. Authors in [14] detail the statistical relational learning on knowledge graphs (KGs) and point out the importance of type constraints and transitivity as well as other statistical patterns or regularities, "which are not universally true but nevertheless have useful predictive power". Similar to [24], they base their method mainly on large scale KBs such as Nell [3], KnowItAll [6], YAGO [22] or DeepDive [20].

The endogenous rule-based inference process has been studied by Zarrouk (2015) and Ramadier (2016) in the framework of RezoJDM, the LSN for French. Their methods rely on the relationships and relationship meta-information that are already present in this LSN in order to propose the new ones following one of the following schemes : deduction, induction (which benefit from taxonomic relations), abduction (exploiting semantic similarity), and inference by refinement. Gelbukh (2018) introduced a comparable inference mechanism to enrich a collocational knowledge base by suggesting new collocations through the inference by abduction (where semantic similarity is calculated on the basis of PWN [7]).

KBs completion can be made using embedding strategies where latent spaces allow modeling candidate facts as resulting from latent factors. RESCAL [15] and TransE [1] propose such approaches. RESCAL performs collective learning using the latent components of the tensor factorization. In other words, the entity neighborhood is used to predict an unknown relation between this entity  $e_1$  and some other entity  $e_2$  knowing that some other entities similar to  $e_1$  (in terms of their neighborhood) are connected to  $e_2$  through the relation type  $t$ . The TransE method models relationships by interpreting them as translations in the embedding space and relies on low-dimensional embeddings of the entities. This system associates some vector depending on the relationship type to the vector of this relationship *tail* (source). This allows learning only one

low-dimensional vector for each entity and each relationship.

Markov Random Fields (MRF) based approach to KBs completion has been proposed with first-order logic representation through a *Markov logic network* in [17]. MRF-based approach with probabilistic logic representation has been introduced in [2]. Path ranking approaches based on random walk i.e. [5] are also being explored.

### III. CONTEXT

To conduct the experiment, we built a multilingual lexical-semantic network (MLSN) inspired by the RezoJDM for the food domain. At the time of our writing, the MLSN contains 821 781 nodes (terms) and 2 231 197 arcs (relationships). It is a directed, typed, and valuated graph. The MLSN  $k$  sub-graphs correspond to each of the  $k$  languages (English, French, Russian, and Spanish). A specific sub-graph fulfills the role of the *interlingual pivot*. The MLSN nodes may correspond to one of the following types : lexical items (i.e. *garlic*); interlingual items (also called *covering terms*) that are not necessarily labeled in a human readable way; relational items (i.e. relationship reifications such as *salad[r\_has\_part]garlic*); category items modeling categories, parts of speech or other morpho-syntactic features (i.e. *Verb :Present, Noun :AccusativeCase*).

During the MLSN set up and building, the hypothesis introduced in [16] concerning the "non separation" between general and domain specific knowledge has been considered. This hypothesis states that during the semantic analysis of domain specific texts that relies on background knowledge, the presence of general common sense knowledge in addition to the domain specific knowledge improves the performance of such analysis. Thus, general commonsense knowledge information has been integrated into our resource from the existing LSNs i.e. PWN, ConceptNet etc. This integration has been "guided" by the domain specific comparable corpora (96 083 cooking recipes, about 8 300 000 words) as its vocabulary has been used as a seed.

A relation  $r \in R$  is a sextuplet  $r = \langle s, t, type, v, l_s, l_t \rangle$  where  $s$  and  $t$  correspond respectively to the source and the target term of the relation. The relation  $type$  is a typical relation type. It may model different features such as taxonomic and part-whole relations ( $r_{isa}$ ,  $r_{hypo}$ ,  $r_{has\_part}$ ,  $r_{matter}$ ,  $r_{holo}$ ), possible predicate-argument relations (typical object  $r_{object}$ , location  $r_{location}$ , instrument  $r_{instr}$  of an action), "modifier" relations (typical characteristic  $r_{carac}$ , typical manner  $r_{manner}$ ) and more<sup>1</sup>. The relationship valuation  $v$  corresponds to the characteristics of the relation which are its weight, confidence score, and annotation. The relation weight may be negative in order to model noise and keep the information about erroneous relations easy to access programmatically so they could not affect the inference processes. The confidence score is a score attributed to a particular data origin (external resource, inference process). In practice, this feature

1. We also introduced more specific relation types such as  $r_{entailment}$ ,  $r_{cause}$ ,  $r_{telic\_role}$ ,  $r_{incompatible}$ ,  $r_{before}$ ,  $r_{after}$  etc.

is an array as different origins may provide the same relation. The confidence information is provided as an argument to the function that maps from some external knowledge resource to the MLSN. In case of relation calculated by an inference process, it corresponds to the precision evaluated on a sample of candidate relations returned by this process. To *annotate* a relation we add a complementary information that allows qualifying this relation. The labels  $l_s$  and  $l_t$  correspond to the language (sub-graph) labels.

As it has been difficult to set up the pivot using a multilingual embedding (joining multiple spaces, one per language) as well as to avoid pairwise alignment based on combinatorial criteria, the pivot has been started as a natural one using the English edition of DBNary [19]. It incrementally evolves based on sense alignments (obtained through external resources or by inference) between the languages of the MLSN to become interlingual. It can be considered as a union of word senses lexicalized or identified in the languages covered by the MLSN. Such progressive pivot building allows reducing the artificial contrast phenomena defined by [18] as a discriminatory information loss linked to the divergent conceptualization and lexicalization observed in different languages. Even though

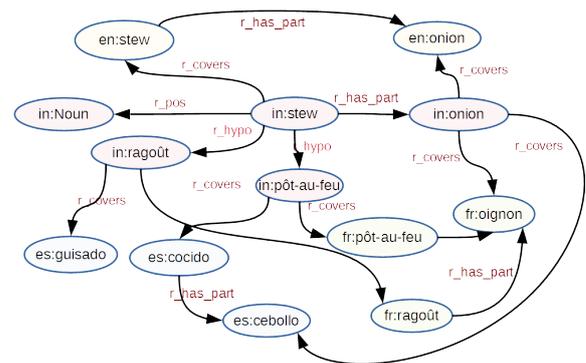


FIGURE 1. Pivoted MLSN architecture. *in* precedes interlingual terms.

we assume the pivot as being interlingual, it is still close to a natural one. As a result, the inference mechanisms we detail are suitable for the architecture by transfer.

The MLSN has been set up and populated through term and relationship extraction from corpora using state of the art pattern based techniques and lexical semantic patterns which add semantic constraints i.e. lookup for semantic relationships in the MLSN to the syntactic pattern; by integrating weakly structured domain specific information (glossaries, vocabulary lists, nutritional composition of food items) as well as structured lexical semantic resources (such as RezoJDM, PWN, ConceptNet) and inference mechanisms.

The table I shows that the global inference impact is higher in the context of the relationship types that are hard to yield by integration from structured resources or identified in corpora. The relationship typed *refinement*( $r_{raff}$ ) connecting terms to their senses will be described in section IV. The pivot coverage (how many terms from a given MLSN sub-graph are connected to the pivot?), the presence of semantic relationships and sense

TABLE I  
MLSN RELATIONSHIP ACQUISITION.

<i>R</i> <sub>type</sub>	corpus	integ.	before inf	inf	prod
<i>r</i> <sub>isa</sub>	67 894	544 632	612 526	27 546	4%
<i>r</i> <sub>hypon</sub>	688	797 783	798 471	41 053	5%
<i>r</i> <sub>has_part</sub>	662 737	172 287	835 024	48 015	6%
<i>r</i> <sub>matter</sub>	606	35 597	36 203	1 893	5%
<i>r</i> <sub>holo</sub>	224	67 081	67 305	51 360	76%
<i>r</i> <sub>object</sub>	42 280	29 262	71 542	15 512	22%
<i>r</i> <sub>carac</sub>	8 300	69 236	77 536	9 521	12%
<i>r</i> <sub>manner</sub>	2 854	3 250	6 104	250	4%
<i>r</i> <sub>location</sub>	2086	3 573	5 659	146	3%
<i>r</i> <sub>instr.</sub>	58	2 738	2 796	402	14%
<i>r</i> <sub>refinement</sub>	221	29 441	29 662	182 135	614%
Overall	788 344	1 754 484	2 542 828	377 816	15%

refinements in a MLSN sub-graph also determine the success of a relationship inference process. Semantic information is easier to obtain from monolingual external resources. Thus, the exogenous data and semantic relationship acquisition are mostly monolingual. As a result, some terms may not be covered by the pivot. As the semantic relationships are used by the inference mechanism for logical filtering, when a MLSN sub-graph has numerous semantic relationships the inference precision is higher.

#### IV. CROSS-LINGUAL SEMANTIC RELATION INFERENCE

**Principle** - In this section we detail the inference of new semantic relations in one lexicalized part of the MLSN from the ones existing in another MLSN part (sub-graph). In a pivoted resource, the relations are first inferred into the pivot (ascending inference). Second, they are inferred in other sub-graphs (descending inference). In transfer-based resources where lexicalized sub-graphs are directly connected to each other, the inference process would directly apply to the source and target languages and rely on translation links between those. Thus, the proposed inference process is independent from the architecture of the resource (transfer or pivot based). It also may be considered as independent from the expressiveness of the multilingual resource as we define for and test it on a very expressive MLSN with numerous relation types. **Monolingual context** - In the monolingual context, the mechanisms of inference by deduction, induction, abduction, and inference with sense refinements apply. These processes have been described in [25]. In case of transitive semantic relations (i.e. hyperonymy, hyponymy), **deduction** and **induction** can be implemented. These inference schemes propose to a term some relevant relations detained by its hypernyms or hyponyms based on the transitivity of these taxonomic relations. For (nearly) synonyms, the **abduction** procedure is chosen. The abduction yields a set of terms similar to the term  $T$  then proposes the neighborhood relations detained by these terms to  $T$ . In order to calculate the similar terms more finely, in addition to calculating Jaccard similarity score, weighted Jaccard such as in [12] or some other similarity measure, we consider semi-relations (Typed ingoing and outgoing relations from/to a neighbor) shared by a pair of similar (synonymous)

terms. **Inference with sense refinements** exploits the sense refinement of polysemous terms. When the senses (we also call *refinements*) are modeled, it is possible to verify whether they are semantically related to the opposite term of the relation to be inferred.

To give an example, we may consider the french term *soupe* and its refinements  $\{soupe>potage, soupe>neige, soupe>repas\}$  ("soup>broth", "soup>melted snow", "soup>meal") and a new candidate relation we want to infer (relation obtained either by deduction, induction or abduction)  $soupe \xrightarrow{r_{carac}} chaud(hot)$ . In order to automatically accept such relation, we may check if one of soup refinements is semantically connected to *chaud* or one of its refinements :  $soupe>potage \xrightarrow{r_{isa}} liquide \& liquide \xrightarrow{r_{carac}} chaud$ .

**Multilingual context** - In the context of the cross-lingual semantic relation inference, we use the *r\_covers* relations to identify the semantic relations that correspond to the premises of the inference rules.

The relations typed *r\_covers* link an interlingual term to the lexicalized terms that it covers. We may suppose that during the ascending ( $language \rightarrow pivot$ ) and the descending ( $pivot \rightarrow language$ ) processes we deal with the equivalent terms. Due to the discrepancies between languages and to the fact that our recent interlingual pivot is still close to the natural one, one lexicalized term may have multiple covering terms and *vice versa*. Therefore, we consider the *r\_covers* relation as a cross-lingual variant of a (possibly) incomplete synonymy. Given that, the case of inference that applies can be either abduction or inference by refinement. For the abduction case in the ascending multilingual context, the relation to be inferred is considered as an abduction rule instance. We transform its source and target terms into the sets which may contain interlingual and lexicalized terms. Then, we explore the neighborhood of the intersection between the obtained sets. If the intersection between the typed semi-relations is sufficient (we empirically set the threshold to 3), the relation from the lexicalized subgraph is proposed for the terms in the interlingual pivot (and *vice versa* while performing a descending inference).

The case of polysemy is processed as if it was an "inference with sense refinements" case. It checks by triangulation the presence of semantic relations between the "refinements" of a term (the different covering or covered terms) and the opposite term of the relation to be inferred. A simplified example of the Russian term *pryanik* for which we are looking to infer relations typed *r\_has\_part* thanks to the "fr" MLSN subgraph illustrates the inference mechanism. The distinction between the sense refinements of *pain d'épices* in French can be modeled at the interlingual level as two interlingual refinements of the interlingual term  $in:gingerbread$  that are  $in:gingerbread>cake$  and  $in:gingerbread>biscuit$ . The inference is a twofold process. The relationships from the "fr" subgraph are inferred into the pivot using the interlingual terms that cover the *pain d'épices* neighbors : such as  $in:sugar \xrightarrow{r_{covers}} fr:sucre$ ,  $in:ginger \xrightarrow{r_{covers}} fr:gingembre$ , etc. Then, the relations are inferred from

the pivot to the "ru" subgraph. As *pryanik* in Russian culinary tradition has the soft cookie texture (this information is available from semantic relations of *pryanik* and from translation links where *pryanik* is linked to both refinements of *pain d'épices*, the distinction observed in French is not relevant for Russian. Thus, the descending inference process proposes candidate relations of the general interlingual term *in:gingerbread* to *pryanik*. As the general term detains the relationships of its refinements, *pryanik* yields all the relationships of *pain d'épices* that can be represented on the interlingual level and persist after logical/statistical filtering.

The abduction scheme generates a lot of candidate relationships. Therefore, a filtering strategy significantly improves the precision. First, we apply part-of-speech pre-filtering can be used depending on relation types. For instance, in the case of the relation typed  $r_{carac}$  (typical characteristic) the source term must be a noun whereas the target term must be an adjective (i.e.  $cake \xrightarrow{r_{carac}} sweet$ ). Second, we use the statistical filtering as the relations of the MLSN can be analyzed in terms of their number, weight, and origin. The weight  $w$  corresponds to the crowd-sourced weight or to the default weight. Similar to ConceptNet, we introduced the information regarding the confidence given to the structured resource from which a given relation has possibly been integrated or to the endogenous inference. Thus we attach the *origin* information to the relationships. It took the form of an array of strings (naming the different processes that provided the relation) to which we associate an array of confidence scores  $\psi = \{i_1, i_2, \dots, i_n\}$  where  $i_j \in [0; 1]$ . The size of the set of semi-relations shared by the terms  $\phi$  is also taken into account. For the positively or negatively weighted relation the filtering function is calculated as follows for  $w \in \mathbb{Z}$  and  $|\psi| > 0$ :

$$f(r) = \phi \times \frac{w}{Max(\psi) \times \log(|\psi|)}$$

In a mature MLSN, the relation inference algorithm becomes a simple lifting and descending algorithm where no significant filtering to be applied.

**Experiment** - We tested our approach on all the semantic relations and languages present in the MLSN. The table II lists the results of the descending inference process. The results are presented in terms of number of relations in the source partition (**#orig**), number of candidate relations (**#cand**), number of accepted relations after filtering (**#acc**), productivity of the algorithm (**prod**), acceptance rate (**%acc**, the percentage candidate relationships that verify the inference rule premises and conclusion and subsist after filtering), and precision (**pr**) which has been manually evaluated on a sample of 500 accepted relations (per type). This type of manual evaluation has been chosen due to the difficulty to find a well balanced reference for evaluation. As we integrated the main LSNs for the languages covered by the MLSN, we presumably infer the relationships that are not explicitly represented in such structured resources. The range **r** has been introduced to express how close a given process is situated to the "gold" productivity (100%). Indeed such "gold" productivity would

mean that the sense based alignment is sufficient for a given term. The table II lists the results for the main semantic relations of the Russian and Spanish sub-graphs and details the evolution of the number of semantic relations.

TABLE II  
DESCENDING INFERENCE OF SEMANTIC RELATIONS.

type	l	#bef	#inf	#aft	ev
$r_{isa}$	ru	46 827	7 036	53 863	+14%
	es	36 807	268 040	304 847	+828%
$r_{has\_p}$	ru	65 772	3 682	69 454	+5%
	es	10 166	56 883	67 049	+559%
$r_{mat}$	ru	5190	4230	9 420	+81%
	es	4013	7 351	7 764	+183%
$r_{man}$	ru	1 265	1 655	2 920	+131%
	es	1 753	9 507	11 260	+542%
$r_{loc}$	ru	640	621	1 261	+97%
	es	90	567	657	+630%
$by\ lang.$	ru	119 694	17 224	136 918	+14%
	es	52 739	342 348	395 087	649%
TOTALS	-	172 433	359 572	532 005	+208%

The logical filtering concerns only a subset of relation types to be checked  $m$  times (according to the branching factor of the term). Thus, the global complexity of the logical filtering would be  $O(m \times n^2)$ . La *average complexity* would correspond to the average degree observed in the MLSN at the time of our writing :  $d_{av} = 4 \Rightarrow O(16 \times m)$ .

**Towards the Sense-based Alignment** - The MLSN refinement relations allow modeling the "use" senses of a term. The refinement corresponds to maximal cliques (calculus) or to the contributions (GWAP). For the french term *baguette*, we distinguish the sense "bread" as opposed to "direction", "stick", and "magic wand". The glossed refinement corresponds to this sense is *baguette > pain*. Thus, we have the following structure in the MLSN :  $baguette \xrightarrow{r_{raff}} baguette > pain \xrightarrow{r_{glose}} pain$ . A glossed refinement may be itself refined and glossed. In the case of a resource that already possesses refinement relations, it is possible to infer some cross-lingual new refinements from the existing ones. The 30% refinement rate of the MLSN pivot has been obtained using this process.

When the term has multiple covering terms, the descending inference pattern can be applied. We consider that the covering terms are potentially linked to the gloss. First, we temporarily label the potential senses using the labels of the covering terms. Second, we group the redundant senses and choose the gloss. the recently started experiment with this pattern allowed producing the first batch of 2 535 sense refinements in Russian whereas 1 800 refinements have been yielded for this language using the glossed refinement technique.

## V. CONCLUSION

We introduced a simple endogenous method for cross-lingual semantic relation inference to improve structured KBs such as MLSN. Given a certain coverage in terms of translation links, it allows enhancing the under-resourced parts of a lexical semantic resource from the rich ones. Even though they benefit from translation resources and tools, some "rare" languages are not covered by any rich lexical semantic resource. To

some extent, the method is beneficial for domain specific MLSNs. It allows rich semantic modeling which provides a semantically structured representation for the fine grained semantic analysis (including word sense disambiguation) and statistical representation learning.

## REFERENCES

- [1] Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States.*, pages 2787–2795, 2013.
- [2] Matthias Bröcheler, Lilyana Mihalkova, and Lise Getoor. Probabilistic similarity logic. In *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence, UAI'10*, pages 73–82, Arlington, Virginia, United States, 2010. AUAI Press.
- [3] Andrew Carlson, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam R. Hruschka Jr., and Tom M. Mitchell. Toward an architecture for never-ending language learning. In *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2010, Atlanta, Georgia, USA, July 11-15, 2010*, 2010.
- [4] Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. Convolutional 2d knowledge graph embeddings. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pages 1811–1818, 2018.
- [5] Xin Dong, Evgeniy Gabrilovich, Jeremy Heitz, Wilko Horn, Ni Lao, Kevin Murphy, Thomas Strohmman, Shaohua Sun, and Wei Zhang. Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14*, pages 601–610, New York, NY, USA, 2014. ACM.
- [6] Oren Etzioni, Michael Cafarella, Doug Downey, Ana-Maria Popescu, Tal Shaked, Stephen Soderland, Daniel S. Weld, and Alexander Yates. Unsupervised named-entity extraction from the web: An experimental study. *Artificial Intelligence*, 165(1):91 – 134, 2005.
- [7] Christiane Fellbaum. *WordNet An Electronic Lexical Database*. The MIT Press, Cambridge, MA ; London, 1998.
- [8] J. Ferber. *Les systèmes multi-agents: vers une intelligence collective*. InterEditions, Paris, 1995.
- [9] Alexander F. Gelbukh. Inferences for enrichment of collocation databases by means of semantic relations. *Computación y Sistemas*, 22(1), 2018.
- [10] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. *KDD : proceedings. International Conference on Knowledge Discovery & Data Mining*, 2016:855–864, 2016.
- [11] Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. A three-way model for collective learning on multi-relational data. In *Proceedings of the 28th International Conference on Machine Learning, ICML'11*, pages 809–816, USA, 2011. Omnipress.
- [12] Jerónimo Hernández-González, Estevam R. Hruschka Jr., and Tom M. Mitchell. Merging knowledge bases in different languages. In *Proceedings of TextGraphs-11: the Workshop on Graph-based Methods for Natural Language Processing*, pages 21–29, Vancouver, Canada, August 2017. Association for Computational Linguistics.
- [13] Mathieu Lafourcade. *Lexique et analyse sémantique de textes - structures, acquisitions, calculs, et jeux de mots. (Lexicon and semantic analysis of texts - structures, acquisition, computation and games with words)*. Montpellier, 2011.
- [14] Mathieu Lafourcade and Lionel Ramadier. Semantic Relation Extraction with Semantic Patterns: Experiment on Radiology Report. In *LREC 2016 Conference on Language Resources and Evaluation*, volume 10th of *LREC 2016 Proceedings*, Portorož, Slovenia, May 2016.
- [15] Maximilian Nickel, Kevin Murphy, Volker Tresp, and Evgeniy Gabrilovich. A review of relational machine learning for knowledge graphs: From multi-relational link prediction to automated knowledge graph construction. *CoRR*, abs/1503.00759, 2015.
- [16] Lionel Ramadier. *Indexation and learning of terms and relations from reports of radiology*. Theses, Université de Montpellier, November 2016.
- [17] Matthew Richardson and Pedro Domingos. Markov logic networks. *Mach. Learn.*, 62(1-2):107–136, February 2006.
- [18] Gilles Sérasset. Dbnary: Wiktionary as a lmf based multilingual rdf network. In *LREC*, 2012.
- [19] Gilles Sérasset. DBnary: Wiktionary as a Lemon-Based Multilingual Lexical Resource in RDF. *Semantic Web – Interoperability, Usability, Applicability*, pages –, 2014. To appear.
- [20] Jaeho Shin, Sen Wu, Feiran Wang, Christopher De Sa, Ce Zhang, and Christopher Ré. Incremental knowledge base construction using deepdive. *Proc. VLDB Endow.*, 8(11):1310–1321, July 2015.
- [21] Robert Speer and Catherine Havasi. Representing general relational knowledge in conceptnet 5. In *LREC Proceedings*, 2012.
- [22] Fabian Suchanek, Gjergji M Kasneci, and Gerhard M Weikum. Yago: A Core of Semantic Knowledge Unifying WordNet and Wikipedia. In *16th international conference on World Wide Web*, Proceedings of the 16th international conference on World Wide Web, pages 697 – 697, Banff, Canada, May 2007.
- [23] Kristina Toutanova and Danqi Chen. Observed versus latent features for knowledge base and text inference. In *Proceedings of the 3rd Workshop on Continuous Vector Space Models and their Compositionality*, pages 57–66, Beijing, China, July 2015. Association for Computational Linguistics.
- [24] Quan Wang, Bin Wang, and Li Guo. Knowledge base completion using embeddings and rules. In *Proceedings of the 24th International Conference on Artificial Intelligence, IJCAI'15*, pages 1859–1865. AAAI Press, 2015.
- [25] Manel Zarrouk, Mathieu Lafourcade, and Alain Joubert. Inference and Reconciliation in a Crowdsourced Lexical-Semantic Network. In *CICLING: International Conference on Intelligent Text Processing and Computational Linguistics*, number 14th, Samos, Greece, March 2013.



# Deep Learning Hyper-parameter Tuning for Sentiment Analysis in Twitter based on Evolutionary Algorithms

Nuria Rodríguez-Barroso, Antonio R. Moya, José A. Fernández, Elena Romero,  
Eugenio Martínez-Cámara, Francisco Herrera

Andalusian Research Institute in Data Science and Computational Intelligence,  
University of Granada,  
Granada (Spain)

Email: {rbnuria,anmomar85,fernandezja,elenaromero}@correo.ugr.es; {emcamara,herrera}@decsai.ugr.es

**Abstract**—The state of the art in Sentiment Analysis is defined by deep learning methods, and currently the research efforts are focused on improving the encoding of underlying contextual information in a sequence of text. However, those neural networks with a higher representation capacity are increasingly more complex, which means that they have more hyper-parameters that have to be defined by hand. We argue that the setting of hyper-parameters may be defined as an optimisation task, we thus claim that evolutionary algorithms may be used to the optimisation of the hyper-parameters of a deep learning method. We propose the use of the evolutionary algorithm SHADE for the optimisation of the configuration of a deep learning model for the task of sentiment analysis in Twitter. We evaluate our proposal in a corpus of Spanish tweets, and the results show that the hyper-parameters found by the evolutionary algorithm enhance the performance of the deep learning method.

## I. INTRODUCTION

OPINIONS, sentiments, experiences, private states, broadly speaking subjective information, are continuously posted on micro-blogging sites as Twitter. The processing of this kind of information is crucial for other users and for any kind of organisation, because it offers a valuable source of knowledge to understand the perspectives of users on topics of interest, which eases the process of making decisions [1]. Sentiment Analysis (SA) is the task centred on labelling the opinion meaning of a text, and it is defined as the computational treatment of opinions, sentiments and subjectivity in texts [2].

Since the use of language in Twitter has its own characteristics that make it different from the use of language in formal genre of writing, specific computational methods have to be developed [3]. The main contributions to the processing of the sentiment of tweets can be found in the respective tasks of the workshops SemEval<sup>1</sup> for the English language and TASS<sup>2</sup> for the Spanish language. The state of the art on those workshops has evolved from the use of linear classification systems grounded in the use of a big bunch of hand-crafted linguistic features [4], [5] to the use of deep learning methods

without the need in most of the cases of hand-crafted features [6], [7].

Besides the strong results of deep learning methods in SA in Twitter, we stress out that those deep learning methods has reduced the need of feature engineering, because they are based on the use of unsupervised pre-train features, which the most used are vectors of word embeddings. Deep learning methods depend on the configuration of some parameters that are known as hyper-parameters, such as the number of output units of each neural layer or the dropout rate. Those hyper-parameters must be defined by hand, hence the positive reduction of the effort in the designing of features has been changed to the effort of setting the right hyper-parameters value. The current trend in the development of neural networks for SA is to attempt to encode as much contextual information as possible, which is the aim, for instance, of the self-attentive networks [8] and memory networks [9]. The high complexity of those deep learning architectures entails to define a higher number of hyper-parameters, which means that their configuration would not be an easy task.

We define the process of hyper-parameter setting as an optimisation task, because the optimisation of the value of the hyper-parameters allows to optimise the performance of the neural network. In this paper we thus claim that the use of an optimisation method, as an Evolutionary Algorithm [10], may find out the right hyper-parameters values and consequently optimise the performance of a neural network. We propose the use of the evolutionary algorithm SHADE [11] for optimising the hyper-parameters of a self-attentive neural network for the task of SA in Twitter.

We evaluate our proposal in the task of SA in Twitter in Spanish, and we used the Spanish set of the corpus InterTASS [12]. We define as a baseline model our self-attentive neural network with a set of hyper-parameters values defined by hand, and we compare its performance with the optimised version of the neural model. Likewise, we compare the performance of our proposal with the results reached in the TASS 2018

<sup>1</sup>[https://aclweb.org/aclwiki/SemEval\\_Portal](https://aclweb.org/aclwiki/SemEval_Portal)

<sup>2</sup><http://www.sepln.org/workshops/tass/>

competition,<sup>3</sup> and we show how our proposal without any external knowledge reaches a similar performance than the highest ranked systems in the competition. Moreover, we show how our evolutionary proposal has the ability to improve the learning of minority classes in a imbalanced dataset, as the InterTASS corpus is, and reduces the complexity of the neural model. Although we evaluated our proposal in an imbalanced dataset, we did not conduct any standard data augmentation technique that are usually performed for enhancing the performance of deep learning methods [13], because our aim is to evaluate our claim without the influence of any data pre-processing method.

The reminder of this paper is organised as what follows: Section II exposes some related works to SA in Twitter and hyper-parameter learning. Subsequently, Section III presents our deep learning model for SA in Twitter, which is optimised by an evolutionary algorithm that is detailed in Section IV. Sections V and VI are focused on the description of the experimental set up and the analysis of the results. Finally, Section VII presents the conclusions of our work.

## II. RELATED WORKS

We propose the automatically learning of the hyper-parameters of a deep learning method in order to tackle the task of SA in Twitter. Accordingly, Section II-A describe some works related to SA, and Section II-B is focused on the task of neural networks hyper-parameters learning.

### A. Sentiment Analysis in Twitter

Since the first days of Twitter, this microblogging site has attracted the attention of the research community, although the first works were closer to social sciences [14] than computer science, as well as to the concept of the electronic word of mouth [15]. However, as the popularity of Twitter was increasing, it was becoming in a communication tool in which users exchange their private states, or in other words their experiences, sentiments and opinions.

The first works on SA in Twitter were similar to the first ones in regular texts [16], [17], they were focused on the study of how to represent the opinion meaning of texts and the comparison of linear machine learning classification algorithms. In [18], the first corpus of SA in Twitter is described, and the authors evaluated the performance of three linear classification methods with three different feature vector representations approaches. The following works centred the efforts on feature engineering, broadly speaking, on the use of linguistic features and external knowledge for the representation of the opinion meaning of tweets. For instance, in [19] the tweets were represented with a combination of weighted unigrams and features generated from a sentiment lexicon. Similarly, in [20] the authors used a list of subjective hashtags besides the use of a sentiment lexicon and unigrams to classify the polarity of tweets from different topics. The use of sentiment external knowledge is essential in [21], in which the authors

first represented the tweets as bag of unigrams and bigrams, and each unigram and bigram is represented as a vector of sentiment values aggregated from several sentiment lexicons.

The classification of the polarity of tweets was also used to the prediction of future events, such as the outcome of elections [22], [23]. Likewise, in [24], the authors use the classification of the opinion to predict the evolution of stock markets. As the previous works, the method are based on the representation of the tweets with a great bunch of features and the use of linear classifiers.

Besides the strong results of deep learning methods in Twitter SA, they allow to extremely reduce the efforts in feature engineering and in the use of external knowledge. However, this is caused by the representation of the input sequences of text, in this case, tweets, with unsupervised pre-trained feature vectors. Those feature vectors are known as word embedding that represent the meaning of each word, and they are based on the distributional semantics hypothesis. Accordingly, deep learning methods allow to reach strong results with a low designing effort. For instance, in [25], the authors classify the polarity and the language of tweets with a convolutional neural network (CNN). Likewise, the straightforward neural network described in [26] also reached good results in SA in Twitter in Spanish. Other example of the use of deep learning methods for SA in texts different from English can be read in [27]. However, in some cases, the enhancing of the performance of polarity classification in Twitter forces to use deeper and more complex deep learning methods. In [28], the authors propose the combination of a Long-Short Term Memory (LSTM) Recurrent Neural network (RNN) layer and CNN layer for polarity classification of tweets written in English.

### B. Hyper-parameters Learning

The trend in SA in Twitter is the addition of more encoding layers (CNN, LSTM), and other kind of mechanisms to increase the capacity of the network to represent the contextual information of the input sequence of text. Those layers depend on a set of configuration parameters or hyper-parameters, which their right definition is essential for the global performance of the neural network. Moreover, regularisation layers, as Dropout or penalty rates for the loss function, are key elements of the architecture of neural networks in order to avoid the over-fitting. Consequently, the design of a neural network required of an effort of selecting the right hyper-parameters for each of the layers of the architecture. Therefore, the feature engineering effort has evolved to hyper-parameter engineering.

The definition of the right hyper-parameters is not an easy task, and there is not any rule of thumb to do it. However, there exist some strategies to address it, as well as, some computational approaches, which we indicate as what follows:

- 1) Brute force. It consists in the exhaustive evaluation of all possible values of all the hyper-parameters, which is not feasible because of limitation of time and computational resources.

<sup>3</sup><http://www.sepln.org/workshops/tass/2018/>

- 2) Grid search. It is a brute force approach constrained by a pre-defined set of hyper-parameters values. This is a feasible strategy because the number of evaluations is lower in comparison with the brute force, and it allows to reach good results as show in [29]. However, the hyper-parameters values must be defined by hand.
- 3) Random search. In [30] is shown that the random search of the values of the hyper-parameters allows the neural network to reach good results. However, the random search cannot assure to find out the values that optimise the performance of the network.
- 4) Bayesian approximations [31]. The positive side of this strategies is that they do not have to completely run the neural network to optimise it, because they are grounded in a approximation. However, the complexity of those methods make them close to be unfeasible and difficult to be parallelised.
- 5) Evolutionary algorithms [10]. As the bayesian approximations, these algorithms seek in the hyper-parater values search space those ones that may optimise the performance of the network. Nevertheless, the own definition of evolutionary algorithms has specific strategies for finding the right values in the search space. Moreover, these algorithms are parallelisable in contrast to bayesian approximations, indeed they are parallelisable in GPUs [32]. In [33] is described the use of the CMA-ES [34] for tuning the hyper-parameters of a neural network. In [35] is again used the CMA-ES algorithm for the otpimisation of a neural network, but in this case for the generation of a language model. The use of evolutionary methods for hyper-parameter tuning has not ceased, and recently in [36] a new evolutionary method has been proposed with positive results.

Since evolutionary algorithms are showing a positive performance on the task of hyper-parameter optimisation, we select that strategy for our experimentation, and we propose the use of the algorighm SHADE for the tuning of the hyper-parameters of a self-attentive neural network for the task of SA in Twitter.

### III. DEEP LEARNING MODEL FOR SA

Since our aim is to show the suitability of evolutionary algorithms for tuning the value of hyper-parameters, we propose a deep neural network with several layers with the aim of encoding as much contextual information as possible, which also goes in the line of the proposals of the state of the art (see Section II-A). In the subsequent sections we describe the architecture of our neural network that is composed of three main layers: (1) encoding layer (see Section III-A), self-attention layer (see Section III-B) and classification layer (see Section III-C).

#### A. Encoding layer

Two kind of information may be encoded from a sequence of text: local and temporal. The local information is the underlying one from the inter-dependencies among words in

a local context. On the other hand, the entire sequence of text has also their own meaning which depends on the relation of all the words. Because of these two kind of information, we define an encoding layer composed of a CNN, focused on the local information, and an RNN LSTM layer, centred on the temporal information.

a) *CNN*: We choose a CNN layer in order to focus on the local information motivated by its sparse interactions and the ability to combine features of a local context. CNNs get this ability by implementing the discrete convolution operator (see Equation 1).

$$s(t) = (x * w)(t) = \sum_{a=-\infty}^{\infty} x(a)w(t-a) \quad (1)$$

where  $x$  is the input and  $w$  the kernel. The output is sometimes referred to as the feature map of size  $\text{CNN}_{fm}$ .

The input of a CNN layer is always a grid-structured dataset. For example, the sequence of vectors  $\mathbf{w} = (w_1, w_2, \dots, w_n)$ . This layer performs the convolution function for a fixed kernel size  $\mathbf{k}$ . For an 1-dimensional CNN, the output is another grid-structured dataset of size  $n \times \text{CNN}_{fm}$ . Equation 2 summarise the definition for an 1-dimensional CNN layer:

$$\begin{aligned} \text{CNN}(\mathbf{w}_{1:n}, \mathbf{k}) &= \mathbf{y}_{1:n} \\ \mathbf{y}_i &= s(\mathbf{k}) \\ \mathbf{w}_i &\in \mathbb{R}^d, \mathbf{k} \in [1, 2, \dots, n] \end{aligned} \quad (2)$$

b) *Bidirectional Long-Short Term Memory*: The election of RNN to capture temporal information is due to the fact that they maintain memory based on information history. These networks are defined by a non-linear function  $\sigma$  applied recursively on a sequence of inputs  $(w_1, w_2, \dots, w_n)$ . The input of  $\sigma$  is a state vector  $\mathbf{s}_{i-1}$  and an element of the sequence input  $\mathbf{w}_i$ . The output of the non-linear function  $\sigma$  is a new state vector  $\mathbf{s}_i$ , which is transformed to the output vector  $y_i$  by a deterministic function  $O$ . Equation 3 summarise the definition:

$$\begin{aligned} \text{RNN}(\mathbf{w}_{1:n}, \mathbf{s}_0) &= \mathbf{y}_{1:n} \\ \mathbf{y}_i &= O(\mathbf{s}_i) \\ \mathbf{s}_i &= R(\mathbf{w}_i, \mathbf{s}_{i-1}); \\ \mathbf{w}_i &\in \mathbb{R}^d, \mathbf{s}_i \in \mathbb{R}^{f(h_{lstm})}, \mathbf{y}_i \in \mathbb{R}^{h_{lstm}} \end{aligned} \quad (3)$$

LSTM is a gating-based architecture of RNN that uses several gates in order to solve the gradient vanishing (or exploding) problem of RNN. However, LSTM still has a limitation, the recurrence is only implemented in one direction (from left to right). Nevertheless, the meaning of each word depends on their surrounding context words, broadly speaking, the words in their left and right. Accordingly, we use a bidirectional LSTM (biLSTM). These networks consist in two consecutive LSTM layers, each one in one direction (forward ( $\text{LSTM}^f$ ) and backward ( $\text{LSTM}^b$ )), encoding the full context information. We formally define biLSTM in Equation 4.

$$\text{biLSTM}(\mathbf{w}_{1:n}) = [\text{LSTM}^f(\mathbf{w}_{1:n}, s_0^f), \text{LSTM}^b(\mathbf{w}_{1:n}, s_0^b)] \quad (4)$$

### B. Self-Attention mechanism

The aim of attention mechanisms is to give the neural network the capacity of selecting what to learn from the input data, as humans do. Attention mechanisms have become an essential part of sequence modelling in a wide range of tasks. They are commonly used in conjunction with a RNN.

The attention mechanism in NLP tasks allow to learn what words are the most salient for the global meaning of a sequence of text, but it does not take into account the dependencies that each word has with the others. Self-Attention mechanism [37] calculates the relation of each word with the others, hence it uses more information in order to identify the most salient words. Since Self-Attention allows to use more contextual information of the input data, we chose it in order to automatically learn the set of more prominent words for the polarity classification of the input tweets.

The input of the attention mechanism is a matrix of features, in our case the output of a dense layer immediately after the biLSTM layer,  $H = (\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_n)$  where  $h_i \in \mathbb{R}^d$ . This mechanism aims at selecting the best linear combination of the  $n$  hidden vectors in  $H$ . The output of the attention layer is a vector of weights  $\mathbf{a}$ , which are calculated according to Equation 5.

$$\mathbf{a} = \text{sigmoid}(\mathbf{w}_{s2} \tanh(\mathbf{W}_{s1} H^T)) \quad (5)$$

where  $\mathbf{W}_{s1}$  is a matrix of size  $c \times d$  and  $\mathbf{w}_{s2}$  a vector of size  $c$ , with  $c$  arbitrary fixed in  $[1, n]$  (usually equal to  $n$ ). The *sigmoid* function ensures that the output weights are in  $[0, 1]$ .

The output of the attention mechanism has to be combined with the processing pipeline in order to select the most salient words from the input. Accordingly, the output of the attention layer is added up to the output of the dense layer that is immediately after the biLSTM layer.

An extension of the mechanism that performs multiple hops of attention can be used. It is enough to replace the vector  $\mathbf{w}_{s2}$  with a matrix  $\mathbf{W}_{s2}$  of dimension  $r \times n$ , where  $r$  is the number of weighted outputs we want to generate.

$$\mathbf{A} = \text{sigmoid}(\mathbf{W}_{s2} \tanh(\mathbf{W}_{s1} H^T)) \quad (6)$$

Finally, the mechanism encodes the weighted sums by multiplying the matrix  $\mathbf{A}$  and the matrix of features  $\mathbf{H}$ , resulting the matrix  $\mathbf{M} = \mathbf{A}\mathbf{H}$ . To ensure the matrix  $\mathbf{M}$  does not suffer from redundancy problems, the mechanism uses a penalisation term in order to encourage the diversity of the weighted sums across different hops of attention.

### C. Classification layer

The classification starts with the tokenization of the sequence of input text ( $n$ ). The meaning of each word is represented with its corresponding word embedding vector,

which is looked up in a set of pre-trained word embeddings vector. Accordingly, the output of the input layer is  $I_{n \times d}$ .

The output of the input layer is processed by the encoding layer. First, a CNN layer of kernel 2 with feature map of size  $\text{CNN}_{fm}$ . Subsequently, we use an one-dimensional maxpooling operation with pool size as two using padding in order to keep the sentence size. Likewise, we add a dropout layer with rate  $dr^1$  after the maxpooling layer.

After the convolution, we use a biLSTM layer with  $h_{lstm}$  hidden units. We use the L2 kernel regulariser with rate  $L_2r^1$  in each LSTM layer. After that, we reduce the dimension of the output of the biLSTM with a dense layer with  $h^1$  hidden units.

We apply the self-attention mechanism at this point in order to capture the relevance of each word with the generated features. We merge the results of the attention mechanism with the previous output by an addition. We apply a fully-connected layer with output size  $n \times h^1$ .

$$\begin{aligned} \text{sigmoid}(y_{n \times h_2}^9) &= \text{pred}_4 \\ \text{Dropout}(y_{n \times h_1}^8, d_r^3) &= y_{n \times h_2}^9 \\ \text{Dense}(y_{n \times h_1}^7) &= y_{n \times h_2}^8 \\ \text{Attention}(y_{n \times h_1}^6) &= y_{n \times h_1}^7 \\ \text{Dropout}(y_{n \times h_1}^5, d_r^2) &= y_{n \times h_1}^6 \\ \text{Dense}(y_{n \times 2h_{lstm}}^4) &= y_{n \times h_1}^5 \\ \text{biLSTM}(y_{n \times \text{CNN}_{fm}}^3) &= y_{n \times 2h_{lstm}}^4 \\ \text{Dropout}(y_{n \times \text{CNN}_{fm}}^2, d_r^1) &= y_{n \times \text{CNN}_{fm}}^3 \\ \text{MaxPooling}(y_{n \times \text{CNN}_{fm}}^1) &= y_{n \times \text{CNN}_{fm}}^2 \\ \text{CNN}(\mathbf{w}_{n \times d}, 2) &= y_{n \times \text{CNN}_{fm}}^1 \end{aligned} \quad (7)$$

Finally, we use two dense layers. The first one with  $h^2$  hidden units, and the second one matching the number of labels with *sigmoid* activation function.<sup>4</sup>

We show the architecture of our model in Figure 1. Furthermore, we summarise the formal definition in Equation 7.

## IV. EVOLUTIONARY OPTIMISATION OF HYPER-PARAMETERS

The increasing complexity of deep learning models in SA are becoming harder the right configuration of the hyper-parameter values of the neural networks. In this section we describe our proposal grounded in the use of a evolutionary algorithm for tuning the hyper-parameters of a deep learning method.

Evolutionary algorithms (EA) are based on the natural evolution of species, which allows to keep promising individuals, that is, best solutions to our problem. The main steps of these types of algorithms are: (1) Initialisation of a random population, (2) evaluation of the population, (3) selection of the parents, (4) crossover and mutation and (5) replacement of

<sup>4</sup>We decided to use the sigmoid function instead of softmax as activation function of the last layer because the sigmoid function reached better results in previous experiments.

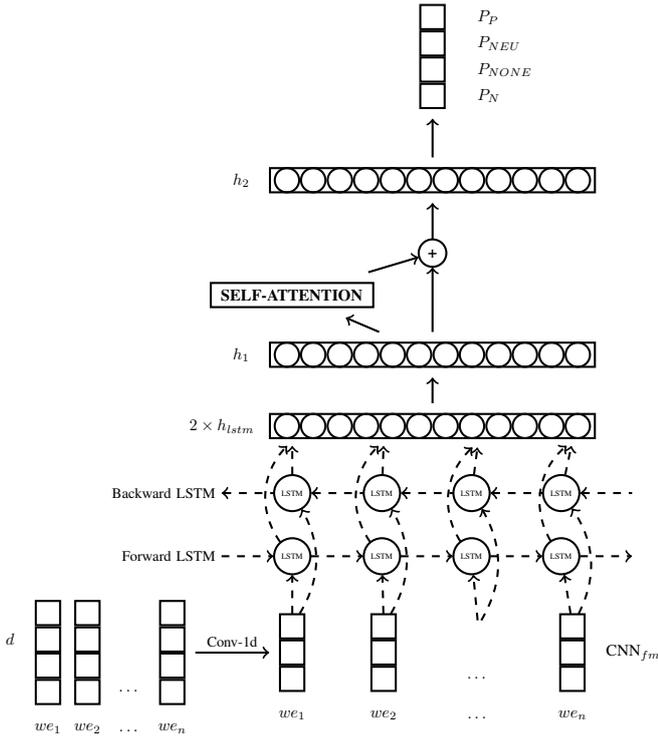


Fig. 1. Architecture of the deep learning model.

the current population by a new generation with individuals selected between parents and offspring. The algorithm is iteratively run until the stopping condition is satisfied.

According to the CEC competition,<sup>5</sup> L-SHADE [38] is in the state of the art of EAs. This algorithm is able to work with large populations of individuals, and it has a mechanism to linearly reduce the population size. The population size of our problem is not large, hence the L-SHADE algorithm is not the most suitable. Consequently, we used the SHADE algorithm [11], which lacks of the linear population size reduction method.

We define as the population of EA the hyper-parameters of some of the layers of our deep learning configuration proposed, specifically: (1) dropout rate ( $dr^1$ ,  $dr^2$  and  $dr^3$ ), (2) regularisation rate (L2) ( $L_2r^1$  and  $L_2r^2$ ) and (3) number of units in the network layers ( $CNN_{f_m}$ ,  $h_{LSTM}$ ,  $h^1$  and  $h^2$ ). Thus, each individual shows a candidate combination of these network hyper-parameters. Figure 2 shows an example of an individual of the population.

110	0.5	64	64	0.0005	0.5	32	0.002	0.5
$CNN_{f_m}$	$dr^1$	$h_{LSTM}$	$h^1$	$L_2r^1$	$dr^2$	$h^2$	$L_2r^2$	$dr^3$

Fig. 2. Example of an individual of the population.

Differential evolution (DE) evolves a population of  $NP$   $D$ -dimensional individual vectors towards the global optimum. We represent as  $x_{i,G}$  the individual  $i$  at generation  $G$  and called it target vector. The initial population should ideally cover the entire search space by randomly distributing each parameter of an individual vector with uniform distribution between prescribed upper and lower parameters bounds.

The main operations of the SHADE algorithm are described as what follows.

#### A. Mutation Operation

Trying to generate diversity in our population, we create a new population which will be crossed in a next step with the current individuals. We define the next mutation operation for this task. For each target vector  $x_{i,G}$  we generate a mutant vector  $v_{i,G}$ . We use the mutation strategy DE/current-to-best/1, which generates a mutant vector using differences between the target vector and the best individual and other random individuals of the current population (see Equation 8).

$$v_{i,G} = x_{i,G} + F(x_{best,G} - x_{i,G}) + F(x_{r_1^i,G} - x_{r_2^i,G}) \quad (8)$$

where the subscripts  $r_1$  and  $r_2$  are random and mutually different integers generated in the range  $[1, NP]$ ,  $F$  is a positive factor for scaling differential vectors and  $x_{best,G}$  is the individual vector with best fitness value in the population at generation  $G$ .

#### B. Crossover Operation

The idea of diversity is needed for seeking the solution. However, we need a balance between exploration of the search-space and exploitation of the current population. Thus, after the mutation operation, crossover operation is used on the individual  $x_{i,G}$  and its corresponding mutant vector  $v_{i,G}$  to generate a trial vector  $u_{i,G}$ , which could be seen as a new individual that allows to keep both properties noted before. For each parameter of the trial vector, we choose between the corresponding parameter of  $x_{i,G}$  or  $v_{i,G}$  depending on crossover rate ( $CR$ ):

$$u_{i,j,G} = \begin{cases} v_{i,j,G} & j = K \text{ or } rand_{i,j}[0, 1] \leq CR \\ x_{i,j,G} & \text{otherwise} \end{cases}$$

where  $CR$  is a value within the range  $[0,1]$ ,  $K$  is a randomly chosen integer in the range  $[1, D]$ . To ensure that the trial vector  $u_{i,G}$  will differ from its corresponding vector  $x_{i,G}$  we add the condition  $j = K$ . As result, we obtain the off-spring population.

#### C. Selection Operation

It selects the best individuals from the population in order to generate a better offspring. The objective function value of each trial vector is compared to its corresponding target vector in the current population. If the trial vector improve the objective function value, the trial vector will replace the target

<sup>5</sup>[http://www3.ntu.edu.sg/home/EPNSugan/index\\_files/cec-benchmarking.htm](http://www3.ntu.edu.sg/home/EPNSugan/index_files/cec-benchmarking.htm)

vector for the next generation. Otherwise, the target vector will remain in the population for the next generation. The selection operation is grounded in the Equation 9

$$x_{i,G+1} = \begin{cases} x_{i,G} & f(x_{i,G}) \leq f(u_{i,G}) \\ u_{i,G} & \text{otherwise} \end{cases} \quad (9)$$

#### D. Parameters self-adaption

The performance of the original DE algorithm is highly dependent on the parameters settings ( $CR$  and  $F$ ). It may require a huge amount of computation time. SHADE can automatically adapt the parameters settings during evolution. For this purpose, SHADE introduce success and failure memories to store different values of  $F$  and  $CR$  within a fixed number of previous generations, hereby named learning period (LP). After the initial LP generations, the probabilities of choosing different parameters values is given by Equation 10.

$$p_{k,G} = \frac{S_{k,G}}{\sum_{k=1}^K S_{k,G}} \quad (10)$$

where  $K$  is the total number of values that we can choose, and  $S_{k,G}$  represents the success rate of the trial vectors generated by the  $k_{th}$  value and successfully entering the next generation within the previous LP generations with respect to generation  $G$ . Equation 11 defines  $S_{k,G}$ .

$$S_{k,G} = \frac{\sum_{g=G-LP}^{G-1} n_{s_{k,G}}}{\sum_{g=G-LP}^{G-1} n_{s_{k,G}} + \sum_{g=G-LP}^{G-1} n_{f_{k,G}}} + \epsilon \quad (11)$$

where  $n_s$  and  $n_f$  are the successful and failures for a certain value in a certain generation.

#### E. Restart mechanism

When an iteration of the evolution is performed, it is possible that our solutions may get stuck in a local search space. Accordingly, we propose to use a restart mechanism in order to avoid to reach a local optimum. When many generations pass without an improvement of the best solution, we opt to restart the population, keeping the best so far. It allows to move our search to new points of the search-space and and test new solutions that could not be evaluated without this approach.

#### F. Objective function

We need an objective function for evaluating the candidate solutions and select best ones. For that, we design a fitness function based on the following requirements:

- We fixed a model (same for each individual).
- Given an individual, each individual parameter is placed adequately in this model.
- We train the model for those values.
- We get predictions and calculate **Macro-F1** (see Section V-C) over a training set.
- With the purpose of minimising the previous value, we use  $1 - MacroF1$  as fitness function for this individual.

Thus, we lead on the evolution of the population towards to the solution with the best results for **Macro-F1** over the evaluation set.

## V. EXPERIMENTS

In this section we show the experiments carried out with our proposed deep learning hyper-parameter tuning based on an EA. We first introduce the dataset used in our experiments, InterTASS Corpus (see Section V-A). Subsequently, we detail the set of pre-train vector of word embeddings used to represent the input tweets (see Section V-B). Then, we compare the results obtained with our method to the ones given by the neural network using the hyper-parameters defined by hand. We also compare our models to the highest ranked model in Task 1 of TASS-2018 Workshop (see Section V-C).

#### A. InterTASS Corpus

The InterTASS Corpus was presented in the *TASS-2018 Workshop* for Task 1, polarity classification at tweet level. The sentiment of the tweets of the corpus are annotated in a scale of 4 levels of polarity intensity: positive (P), Negative (N), neutral (NEU) and no opinion (NONE). The InterTASS Corpus is divided into three datasets: Training (1008 tweets), Development (506 tweets) and Test (1899 tweets). The distribution among the different labels is shown in Table I.

TABLE I  
SIZE OF EACH CLASS IN EACH SUBSET OF THE INTERTASS CORPUS.

	Training	Dev.	Test
P	317	156	642
NEU	133	69	216
N	416	219	767
NONE	138	62	274

According to Table I, the size of the training set is not large, and the distribution of the classes is not balanced, because there is a big difference among the classes P and N and the classes N and NONE. Thus, this two facts will make harder the classification and the optimisation of the model. According to [39], the imbalanced of the data in machine learning may be smoothed by oversampling the minority class. Hence, we slightly reduced the imbalance of the corpus conducting an oversampling method, which consisted in duplicating the instances from the two minority classes. The distribution among the classes in the training set after the oversampling is shown in Table II.

TABLE II  
DISTRIBUTION OF LABELS AFTER OVERSAMPLING THE MINORITY LABELS.

	No Oversample	Oversample
P	317	317
NEU	133	266
N	416	416
NONE	138	276

As we can see from the distribution of classes in table II, NEU and NONE are still the minority classes. However,

the difference with the majority classes has decreased. Establishing the percentage of oversampling is a difficult task, since the amount of data from the minority classes must be increased without losing the representativity of the dataset. For that reason, we choose low oversampling ratios.

### B. Word Embeddings

As we indicated in Section III-A, each word is represented with a vector from a set of pre-trained set of word embeddings. Since the aim is to classify data from Twitter, we trained the embeddings on a set of tweets written in Spanish<sup>6</sup> and using the FastText method [40]. This set of embeddings have a vector representation for some meta-tokens of Twitter, such as: mentions (@user), emojis<sup>7</sup> and for the hashtags of the embeddings training set.

The dimension of the vectors given by these embeddings is  $d = 100$ . Since we used TensorFlow for developing our deep learning method, we had to define a fixed size for the input of the neural network, and to used a zero-padding approach for those tweets shorter and larger to the pre-defined size. The longest tweet in the training set contained 35 tokens, so shorter tweets were filled using padding and truncated in the case of finding a longer tweet in the validation or test set. As the length of the embeddings is 100 and the length of the tweets was set to be 35, the input of the model is a  $35 \times 100$  matrix.

### C. Results

In this section we present the results of the evaluation, that was conducted using the standard evaluation measures in text classification tasks, specifically: F1 score and Accuracy. The F1 is the harmonic mean of the Precision and the Recall, and it provides a trade off among them. Since we are facing up a multi-class classification problem, we used the macro-average version of F1.

We define a set of default values for the hyper-parameters of our deep learning model. Those values were used to configure out the model that was not optimised by the EA algorithm, and they were also used as the initial values of the neural model that was optimised. Table III shows the default hyper-parameters values, which are similar of other deep learning models from the state of the art in SA in Twitter. Likewise, Table III shows the hyper-parameter values returned by the EA algorithm. Some of those values are far from the default ones, and we highlight the value for the second layer of dropout ( $dr^2$ ) that is a very uncommon rate value for a dropout layer, which is usually about 0.5. We also stress out the value for the output units of the biLSTM layer, which is far away from the default value, and it significantly reduces the number of trained parameters of the neural network. Likewise, the size of the output dimension of the CNN was also shortened. Consequently, the SHADE algorithm also optimised the complexity of the neural network.

<sup>6</sup>The tweets to train the set of word embeddings are totally different from the tweets of the training set of InterTASS corpus.

<sup>7</sup>There is an embedding vector for each emoji.

TABLE III  
HYPER-PARAMETER VALUE BEFORE AND AFTER USING EVOLUTIONARY ALGORITHM.

	Starting point	After tuning
CNN $f_m$	128	108
$h_{lstm}$	64	28
$h^1$	32	21
$h^2$	16	21
$dr^1$	0.35	0.471887870
$dr^2$	0.35	0.0706515485
$dr^3$	0.5	0.509543630
$L_2 r^1$	0.0001	0.000410222222
$L_2 r^2$	0.001	0.00173633267

The objective function of the SHADE algorithm was configured out to optimise the F1 score on the validation set. Table IV shows the results reached by the non-optimised deep learning method, our baseline, and the optimised model.

TABLE IV  
RESULTS OBTAINED WITH THE DIFFERENT MODELS.

	Macro-F1	Accuracy
Baseline Model	0.41870	0.60242
Our proposal	0.48352	0.56398

According to Table IV, there is an improvement of more than 6 points in the Macro-F1 after tuning the hyper-parameters with the SHADE algorithm. The use of evolutionary algorithms to tune the hyper-parameters proves to be successful as it improves the Macro-F1 of the initial model. However, the value of the Accuracy in the optimised model is slightly lower than the one reached by the baseline. This is an expected behaviour because of the imbalanced nature of the data. Although the total number of true positives in all the classes is slightly lower in the optimised model, the trade off of correctly tweets classified in all the classes is better in the optimised model as we show in Section VI.

Finally, we use McNemar statistical test [41] in order to study if there are significant differences among the non-optimised model and the optimised one. The test returned that our proposal is significantly better with a  $p$ -value of 0.001 ( $p < 0.001$ ).

Table V shows the position of our proposal in the competition TASS 2018. The first two ranked models elirf-es-run-1 [7] and retuyt-lstm-es-1 [42] are based on deep learning methods, but both of them are grounded in the use of data augmentation techniques. Moreover, the elirf-es-run-1 system also uses external knowledge, such as lists of opinion bearing words in order to enrich with sentiment information of the

TABLE V  
RESULTS OF *InterTass-2018 Workshop task 1*.

	Macro-F1	Accuracy
elirf-es-run-1	0.503	0.612
retuyt-lstm-es-1	0.499	0.549
Our proposal	0.484	0.564
atalaya-ubav3-100-3-syn	0.476	0.544
retuyt-svm-es-2	0.473	0.584

vectors of word embeddings. In contrast, our proposal does not use any amount of external knowledge, and it only uses to train the model the training data. Furthermore, we only duplicated the instances of the minimum classes, which is a less sophisticated data augmentation technique than the one used in *retuyt-lstm-es-1*. Nevertheless, the results of our optimised proposal are close to the first ones.

## VI. ANALYSIS

In this section we study the performance of the experiments explained in the previous section in a more exhaustive way.

In order to explain the results obtained in table IV, we compute the F1 score for each class. We aim to analyse the increases and decreases of the F1 score for each class, which shows the behaviour of the evolutionary algorithm in the task of optimising the F1 score. The F1 for each class can be found in Table VI.

TABLE VI  
RESULTS OBTAINED WITH THE DIFFERENT MODELS SHOWING F1 BY CLASS.

	Baseline Model	Our Model
$F1_P$	0.6691	0.6485
$F1_{NEU}$	0	0.1909
$F1_N$	0.6886	0.67452
$F1_{NONE}$	0.3171	0.4202

Regarding the base model, we see a low performance of the two minority labels (NEU and NONE). We highlight that the base model does not classify any tweet as NEU, which means that the model is not able to learn anything about this label. Likewise, the performance on the NONE label is also reduced, which means that the base model is over-fitted to the labels with more instances. The main improvement of the optimised model is that it improves the performance of the classification in the two minority classes, which improves the performance of the overall system. Consequently, the macro-F1 score of the optimised model is higher, as we indicated in Section V-C.

To go further into this analysis, we examine the behaviour of both models in specific tweets of the different classes. On the first place, we observe that there are some tweets of the majority classes (P and N) that the base model labels correctly and the proposed model does not. We show some of these examples in the table VII. We highlight that the proposed model misclassifies the tweets with the minority classes and, it does not misclassifies among the two majority classes (P and N).

In the same way, there are several examples of tweets of the minority classes (NEU and NONE) that the proposed model labels correctly while the base model does not. We show some examples in table VIII.

This analysis explains the behaviour of the Macro-F1 and accuracy in Table IV. The baseline model (non-optimised) labels correctly more instances but ignoring minority classes while the optimised model deals better with imbalance by giving more importance to minority classes. This illustrates

the importance of choosing a good evaluation measure. Depending on the problem there are evaluation measures that are more representative than others. In our problem, the measure Macro-F1 measures the performance of the models in a more representative way since it takes into account the imbalance. Therefore, according to this measure, we can conclude that the optimised model has provided better results for the imbalanced classification problem.

## VII. CONCLUSIONS

In this paper, we have stress out the difficulty of defining the right hyper-parameters of deep learning method, which makes harder as the complexity of the network increases. We claim that evolutionary algorithms may be used to optimise the value of those hyper-parameters, and we thus propose the use of the SHADE algorithm in order to optimise a self-attentive neural network. We evaluate our proposal in the task of SA in Twitter, specifically of tweets written in Spanish from the InterTASS corpus.

The results show how our optimised proposal allows to improve the performance of the global model and the performance on each of the four classes of the dataset. Likewise, the resultant configuration of the neural network has less parameters than the non-optimised, which is also positive in the sense than optimised the efficiency of the model. Therefore, we conclude that evolutionary algorithms, in our case the SHADE algorithm, are suitable for optimising the configuration of neural networks, broadly speaking, for tuning the hyper-parameter values of deep learning methods. Accordingly, this results open a research line for the meta-learning of hyper-parameters and neural networks, where there a lot of room of improvement.

As future work, we will study the performance of evolutionary algorithms for optimising the number of encoding and classification layers. Likewise, we will evaluate the model with data augmentation approaches to study the synergy between both methodologies.

## ACKNOWLEDGEMENTS

This work was supported by proyect TIN2017-89517-P, by the Spanish “Ministerio de Economía y Competitividad”, the project DeepSCOP-Ayudas Fundación BBVA a Equipos de Investigación Científica en Big Data 2018 and a grant from the Fondo Europeo de Desarrollo Regional (FEDER). Eugenio Martínez Cámara was supported by the Spanish Government Programme Juan de la Cierva Formación (FJCI-2016-28353).

## REFERENCES

- [1] D. Zimbra, A. Abbasi, D. Zeng, and H. Chen, “The state-of-the-art in twitter sentiment analysis: A review and benchmark evaluation,” *ACM Trans. Manage. Inf. Syst.*, vol. 9, no. 2, pp. 5:1–5:29, Aug. 2018. doi: 10.1145/3185045. [Online]. Available: <http://doi.acm.org/10.1145/3185045>
- [2] B. Pang and L. Lee, “Opinion mining and sentiment analysis,” *Found. Trends Inf. Retr.*, vol. 2, no. 1-2, pp. 1–135, Jan. 2008. doi: 10.1561/1500000011

TABLE VII

TWEETS OF MAJORITY CLASSES CORRECTLY LABELED BY BASE MODEL. WE SHOW THE ORIGINAL TWEET IN SPANISH AND ITS ENGLISH TRANSLATION.

Original Tweet	Translated Tweet	Label	Label <sub>base</sub>	Label <sub>ev.</sub>
Gracias a toda la gente que dio RT a mi mensaje de buscar clan. He conseguido ser suplente adc en DragonsZGaminG. Con ganas	Thanks to all the people who gave RT to my clan search message. I have managed to be adc substitute at DragonsZGaminG, I am looking forward to it.	P	P	NEU
Está tocando Bolbora en mi pueblo y yo en la cama con 38 de fiebre	Bolbora is playing in my city and I'm in bed with a fever of 38	P	P	NONE
@user lo sé, lo sé ... la sigo desde hace tiempo jajajaja es de mis favoritas	@user I know, I know... I've been following her for a long time jajaja she's one of my favourites	N	N	NEU
@user la experiencia en mi centro es que los docentes (en la medida de nuestras posibilidades) las llevamos a cabo habitualmente	the experience in my school is that we teachers (to the best of our ability) carry them out regularly	N	N	NONE

TABLE VIII

TWEETS OF MINORITY CLASSES CORRECTLY LABELLED BY OUR PROPOSED MODEL. WE SHOW THE ORIGINAL TWEET IN SPANISH AND ITS ENGLISH TRANSLATION.

Original Tweet	Translated Tweet	Label	Label <sub>base</sub>	Label <sub>ev.</sub>
Venga.. en el próximo tweet os muestro el equipo con el que haré el Modo Carrera en FIFA 17	Alright... In the next tweet I'm showing the team I'm using in the Manager Mode in FIFA 17	NONE	P	NONE
@user sí, las classicas de un día es lo que tienen	@user yes, one day's classical is what it means	NONE	P	NONE
@user mejor a 3. El lunes más primeras Impresiones	@user 3 is better. On Monday more first impressions	NONE	P	NONE
¿Me podrían recomendar alguna película antigua que les guste?	Could anyone recommend me any old film that you like?	NONE	P	NONE
@user yo solo puse las evos del eevee	@user I just put the eeve's evolutions	NONE	N	NONE
Al igual solo es estrés, perdón	Maybe it is just stress, sorry @user	NONE	N	NONE
Luego lo más seguro haga periscope, alguien me va a ver?	I will surely be in live in periscope later, is anyone seeing me?	NONE	N	NONE
Estoy sola en un banco	I'm alone in a bench	NONE	N	NONE
Cada vez estoy más Chetado en el LOL	I'm getting more and more cheated in LOL	NEU	P	NEU
@user pero si ya estás pillada	@user but you are already mad	NEU	P	NEU
Soy una atrevida ay todos me lo dicen	I'm daring ay everyone tell me	NEU	P	NEU
@user pero no es lo mismo	@user but it is not the same	NEU	P	NEU
@user de aquí nace una bonita amistad Pero bueno, si la decepción es solo por el software ni tan mal! @user @user	My doubles are improving, more tomorrow. It's a pity that I cannot show the screen steps (by the moment)	NEU	N	NEU
Esta noche hasta el culete	Tonight I'll be off my face	NEU	N	NEU
Mejorando esos dobles, mañana más. Lástima no poder mostrar los steps de pantalla (Por ahora)	My doubles are improving, more tomorrow. It's a pity that I cannot show the screen steps (by the moment)	NEU	N	NEU
a mí me sigues y no soy guapo	You follow me and I'm not handsome	NEU	N	NEU

- [3] E. Martínez-Cámara, M. T. Martín-Valdivia, L. A. Ureña López, and A. Montejo-Ráez, "Sentiment analysis in Twitter," *Natural Language Engineering*, vol. 20, no. 1, p. 1–28, 2014. doi: 10.1017/S1351324912000332
- [4] S. Mohammad, S. Kiritchenko, and X. Zhu, "NRC-canada: Building the state-of-the-art in sentiment analysis of tweets," in *Second Joint Conference on Lexical and Computational Semantics (\*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*. Atlanta, Georgia, USA: Association for Computational Linguistics, Jun. 2013, pp. 321–327. [Online]. Available: <https://www.aclweb.org/anthology/S13-2053>
- [5] L. Hurtado, F. Pla, and D. Buscaldi, "ELiRF-UPV at TASS 2015: Sentiment analysis in twitter," in *Proceedings of TASS 2015: Workshop on Sentiment Analysis at SEPLN co-located with 31st SEPLN Conference (SEPLN 2015)*. Alicante, Spain: Spanish Society for Natural Language Processing, 2015, pp. 75–79.
- [6] M. Cliche, "BB\_twtr at SemEval-2017 task 4: Twitter sentiment analysis with CNNs and LSTMs," in *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*. Vancouver, Canada: Association for Computational Linguistics, Aug. 2017. doi: 10.18653/v1/S17-2094 pp. 573–580. [Online]. Available: <https://www.aclweb.org/anthology/S17-2094>
- [7] H. L. González, José-Ángel and F. Pla, "ELiRF-UPV at TASS 2018: Sentiment analysis in twitter based on deep learning," in *Proceedings of TASS 2018: Workshop on Semantic Analysis at SEPLN (TASS 2018) co-located with 34nd SEPLN Conference (SEPLN 2018)*. Sevilla, Spain: Spanish Society for Natural Language Processing, 2018, pp. 37–44.
- [8] A. Ambartsumian and F. Popowich, "Self-attention: A better building block for sentiment analysis neural network classifiers," in *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*. Brussels, Belgium: Association for Computational Linguistics, Oct. 2018. doi: 10.18653/v1/P17 pp. 130–139. [Online]. Available: <https://www.aclweb.org/anthology/W18-6219>
- [9] N. Majumder, S. Poria, A. Gelbukh, M. S. Akhtar, E. Cambria, and A. Ekbal, "IARM: Inter-aspect relation modeling with memory networks in aspect-based sentiment analysis," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. Brussels, Belgium: Association for Computational Linguistics, Oct.-Nov. 2018, pp. 3402–3411. [Online]. Available: <https://www.aclweb.org/anthology/D18-1377>
- [10] B. Li, J. Li, K. Tang, and X. Yao, "Many-objective evolutionary algorithms: A survey," *ACM Comput. Surv.*, vol. 48, no. 1, pp. 13:1–13:35, Sep. 2015. doi: 10.1145/2792984. [Online]. Available: <http://doi.acm.org/10.1145/2792984>
- [11] R. Tanabe and A. Fukunaga, "Success-history based parameter adaptation for differential evolution," in *2013 IEEE congress on evolutionary computation*. IEEE, 2013. doi: 10.1109/CEC.2013.6557555 pp. 71–78.
- [12] M. C. Díaz-Galiano, M. A. García-Cumbreras, M. García-Vega, Y. Gutiérrez, E. M. Cámara, A. Piad-Morffis, and J. Villena-Román, "TASS 2018: The strength of deep learning in language understanding tasks," *Procesamiento del Lenguaje Natural*, vol. 62, pp. 77–84, 2019. doi: 10.26342/2019-62-9
- [13] S. Tabik, D. Peralta, A. Herrera-Poyatos, and F. Herrera, "A snapshot of image pre-processing for convolutional neural networks: case study of MNIST," *International Journal of Computational Intelligence Systems*, vol. 10, no. 1, pp. 555–568, 2017.
- [14] A. Java, X. Song, T. Finin, and B. Tseng, "Why we twitter:

- Understanding microblogging usage and communities,” in *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis*, ser. WebKDD/SNA-KDD '07. New York, NY, USA: ACM, 2007. doi: 10.1145/1348549.1348556. ISBN 978-1-59593-848-0 pp. 56–65. [Online]. Available: <http://doi.acm.org/10.1145/1348549.1348556>
- [15] B. J. Jansen, M. Zhang, K. Sobel, and A. Chowdury, “Microblogging as online word of mouth branding,” in *CHI '09 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '09. New York, NY, USA: ACM, 2009. doi: 10.1145/1520340.1520584. ISBN 978-1-60558-247-4 pp. 3859–3864. [Online]. Available: <http://doi.acm.org/10.1145/1520340.1520584>
- [16] B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up? sentiment classification using machine learning techniques,” in *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Jul. 2002. doi: 10.3115/1118693.1118704 pp. 79–86. [Online]. Available: <https://www.aclweb.org/anthology/W02-1011>
- [17] P. Turney, “Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews,” in *Proceedings of 40th Annual Meeting of the Association for Computational Linguistics*. Philadelphia, Pennsylvania, USA: Association for Computational Linguistics, Jul. 2002. doi: 10.3115/1073083.1073153 pp. 417–424. [Online]. Available: <https://www.aclweb.org/anthology/P02-1053>
- [18] A. Go, R. Bhayani, and L. Huang, “Twitter sentiment classification using distant supervision,” Stanford University, Stanford, CA, USA, Tech. Rep. CS224N Project Report, 2009.
- [19] L. Jiang, M. Yu, M. Zhou, X. Liu, and T. Zhao, “Target-dependent twitter sentiment classification,” in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*. Portland, Oregon, USA: Association for Computational Linguistics, Jun. 2011, pp. 151–160. [Online]. Available: <https://www.aclweb.org/anthology/P11-1016>
- [20] L. Zhang, R. Ghosh, M. Dekhil, M. Hsu, and B. Liu, “Combining lexicon-based and learning-based methods for twitter sentiment analysis,” HP Laboratories, USA, Tech. Rep. HPL-2011-89, 2011.
- [21] E. Martínez Cámara, M. A. García Cumbreiras, M. T. Martín Valdivia, and L. A. Ureña López, “SINAI-EMMA: Vectors of words for sentiment analysis in twitter,” in *Proceedings of TASS 2015: Workshop on Sentiment Analysis at SEPLN co-located with 31st SEPLN Conference (SEPLN 2015)*. Alicante, Spain: Spanish Society for Natural Language Processing, 2015, pp. 41–46.
- [22] A. Tumasjan, T. Sprenger, P. Sandner, and I. Welpe, “Predicting elections with twitter: What 140 characters reveal about political sentiment,” 2010.
- [23] A. Jungherr, P. Jürgens, and H. Schoen, “Why the pirate party won the german election of 2009 or the trouble with predictions: A response to tumasjan, a., sprenger, t. o., sander, p. g., & welpe, i. m. “predicting elections with twitter: What 140 characters reveal about political sentiment”,” *Social Science Computer Review*, vol. 30, no. 2, pp. 229–234, 2012. doi: 10.1177/0894439311404119
- [24] J. Bollen, H. Mao, and X. Zeng, “Twitter mood predicts the stock market,” *Journal of Computational Science*, vol. 2, no. 1, pp. 1 – 8, 2011. doi: 10.1016/j.joics.2010.12.007
- [25] J. Wehrmann, W. E. Becker, and R. C. Barros, “A multi-task neural network for multilingual sentiment classification and language detection on twitter,” in *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, ser. SAC '18. New York, NY, USA: ACM, 2018. doi: 10.1145/3167132.3167325. ISBN 978-1-4503-5191-1 pp. 1805–1812.
- [26] F. M. Luque and J. M. Pérez, “Atalaya at TASS 2018: Sentiment analysis with tweet embeddings and data augmentation,” in *Proceedings of TASS 2018: Workshop on Sentiment Analysis at SEPLN co-located with 34th SEPLN Conference (SEPLN 2018)*. Sevilla, Spain: Spanish Society for Natural Language Processing, 2018, pp. 29–35.
- [27] J. Kapočūtė-Dzikienė, R. Damaševičius, and M. Woźniak, “Sentiment analysis of lithuanian texts using traditional and deep learning approaches,” *Computers*, vol. 8, no. 1, 2019. doi: 10.3390/computers8010004. [Online]. Available: <https://www.mdpi.com/2073-431X/8/1/4>
- [28] N. Chen and P. Wang, “Advanced combined lstm-cnn model for twitter sentiment analysis,” in *2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS)*, Nov 2018. doi: 10.1109/CCIS.2018.8691381 pp. 684–687.
- [29] Y. Kim, “Convolutional neural networks for sentence classification,” in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, Oct. 2014. doi: 10.3115/v1/D14-1181 pp. 1746–1751. [Online]. Available: <https://www.aclweb.org/anthology/D14-1181>
- [30] J. Bergstra and Y. Bengio, “Random search for hyper-parameter optimization,” *Journal of Machine Learning Research*, vol. 13, pp. 281–305, Feb. 2012.
- [31] J. Snoek, H. Larochelle, and R. P. Adams, “Practical bayesian optimization of machine learning algorithms,” in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 2951–2959.
- [32] A. Cano, A. Zafra, and S. Ventura, “Speeding up the evaluation phase of gp classification algorithms on gpus,” *Soft Computing*, vol. 16, no. 2, pp. 187–202, Feb 2012. doi: 10.1007/s00500-011-0713-4
- [33] I. Loshchilov and F. Hutter, “Cma-es for hyperparameter optimization of deep neural networks,” in *Proceedings of ICLR 2016 - Workshop Track*, 2014, pp. 1746–1751.
- [34] N. Hansen and A. Ostermeier, “Completely derandomized self-adaptation in evolution strategies,” *Evol. Comput.*, vol. 9, no. 2, pp. 159–195, Jun. 2001. doi: 10.1162/106365601750190398
- [35] T. Tanaka, T. Moriya, T. Shinozaki, S. Watanabe, T. Hori, and K. Duh, “Automated structure discovery and parameter tuning of neural network language model based on evolution strategy,” in *2016 IEEE Spoken Language Technology Workshop (SLT)*, Dec 2016. doi: 10.1109/SLT.2016.7846334 pp. 665–671.
- [36] Y. Nalçakan and T. Ensari, “Decision of neural networks hyperparameters with a population-based algorithm,” in *Machine Learning, Optimization, and Data Science*, G. Nicosia, P. Pardalos, G. Giuffrida, R. Umerton, and V. Sciaccia, Eds. Cham: Springer International Publishing, 2019. ISBN 978-3-030-13709-0 pp. 276–281.
- [37] Z. Lin, M. Feng, C. N. dos Santos, M. Yu, B. Xiang, B. Zhou, and Y. Bengio, “A structured self-attentive sentence embedding,” in *International Conference on Learning Representations (ICLR 2017)*, 2017. [Online]. Available: [https://openreview.net/forum?id=BJC\\_jUqxe](https://openreview.net/forum?id=BJC_jUqxe)
- [38] R. Tanabe and A. S. Fukunaga, “Improving the search performance of shade using linear population size reduction,” in *2014 IEEE congress on evolutionary computation (CEC)*. IEEE, 2014. doi: 10.1109/CEC.2014.6900380 pp. 1658–1665.
- [39] A. Fernández, S. García, M. Galar, R. C. Prati, B. Krawczyk, and F. Herrera, *Learning from imbalanced data sets*. Springer, 2018.
- [40] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, “Enriching word vectors with subword information,” *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 135–146, 2017. doi: 10.1162/tacl\_a\_00051. [Online]. Available: [https://doi.org/10.1162/tacl\\_a\\_00051](https://doi.org/10.1162/tacl_a_00051)
- [41] Q. McNemar, “Note on the sampling error of the difference between correlated proportions or percentages,” *Psychometrika*, vol. 12, no. 2, pp. 153–157, Jun 1947. doi: 10.1007/BF02295996. [Online]. Available: <https://doi.org/10.1007/BF02295996>
- [42] L. Chiruzzo and A. Rosá, “RETUYT-InCo at TASS 2018: Sentiment analysis in spanish variants using neural networks and svm,” in *Proceedings of TASS 2018: Workshop on Sentiment Analysis at SEPLN co-located with 34th SEPLN Conference (SEPLN 2018)*. Sevilla, Spain: Spanish Society for Natural Language Processing, 2018, pp. 57–63.

# Knowledge Extraction and Applications utilizing Context Data in Knowledge Graphs

Jens Dörpinghaus\*, Andreas Stefan†

Fraunhofer Institute for Algorithms and Scientific Computing,  
Schloss Birlinghoven, Sankt Augustin, Germany

Email: \*jens.doerpinghaus@scai.fraunhofer.de, †andreas.stefan@scai.fraunhofer.de

**Abstract**—Context is widely considered for NLP and knowledge discovery since it highly influences the exact meaning of natural language. The scientific challenge is not only to extract such context data, but also to store this data for further NLP approaches. Here, we propose a multiple step knowledge graph-based approach to utilize context data for NLP and knowledge expression and extraction. We introduce the graph-theoretic foundation for a general context concept within semantic networks and show a proof-of-concept-based on biomedical literature and text mining. We discuss the impact of this novel approach on text analysis, various forms of text recognition and knowledge extraction and retrieval.

CONTEXT is a widely discussed topic in text mining and knowledge extraction since it is highly relevant to mine the semantic correct sense of unstructured text. For example in [1], Nenkova and McKeown discuss the influence of context on text summarization. Ambiguity does not only appear for common language words, but especially in scientific context. The scientific challenge is not only to extract such context data, but also to store this data for further NLP approaches. Here, we propose a multiple step knowledge graph-based approach to utilize context data for NLP and knowledge expression. We present a proof of concept based on biomedical literature and show an outlook on further improvements towards next generation knowledge extraction for example for training approaches from artificial intelligence and machine learning.

Knowledge graphs play in general an important role in recent knowledge mining and discovery. A *knowledge graph* (sometimes also called a *semantic network*) is a systematic way to connect information and data to knowledge on a more abstract level than language graphs. It is thus a crucial concept on the way to generate knowledge and wisdom, to search within data, information and knowledge. The context is a significant topic to generate knowledge or even wisdom. Thus, connecting knowledge graphs with context is a crucial feature.

Here, we use a quite general definition of context data. We assume that every information entity can also be a context information for other entities. For example a document can also be a context for other documents (e.g. by citing or referring to the other publication). An author is both a meta information to a document, but also itself context (by other publications, affiliations, co-author networks, ...). Other data is more obvious a context: named entities, topic maps, keywords, etc. extracted with text mining from documents. But already

relations extracted from a text may stand for themselves, occurring in multiple documents and still valuable without the original textual information.

Starting with a simple document graph, in a first step we add context meta information, see figure 1. This will lead to a first knowledge graph which can be used for a first context-based text mining approach. The text mining approach will add more context data, for example from ontologies or relations extracted from the text. The graph with the additional context data can be used as starting basis for more detailed text mining approaches utilizing the novel context data. This step can be redone several time.

In addition using a graph structure has several more advantages for knowledge extraction in biological and medical research. Here scientists are for example interested in exploring the mechanisms of living organisms and gaining a better understanding of underlying fundamental biological processes of life. Today the biomedical field mostly relies on systems biology approaches such as integrative knowledge graphs to decipher mechanism of a disease, by considering system as a whole which is considered as a holistic approach. In that, disease modeling and pathway databases play an important role. Knowledge Graphs built using Biological Expression Language (BEL, see [www.openbel.org](http://www.openbel.org)) is widely applied in biomedical domain to convert unstructured textual knowledge into a computable form. The BEL statements that form knowledge graphs are semantic triples that consist of concepts, functions and relationships [2]. In addition, several databases and ontologies implicitly form a Knowledge Graph. For example Gene Ontology, see [3] or DrugBank, see [4] or [5] cover a huge amount of relations and references to other fields.

Over the last few years new domain specific languages (DSL) and knowledge representations like BEL [6] have been proposed to publish and store this kind of statements and findings. There are still several crucial issues converting literature to knowledge. For example the quality and completeness of such networks has to be evaluated. And with this, to generate new knowledge the context of concepts in a Knowledge Graph has to be considered.

We will first of all give a preliminary overview about information theory and management. With this, we will introduce and discuss the novel approach of managing and mining

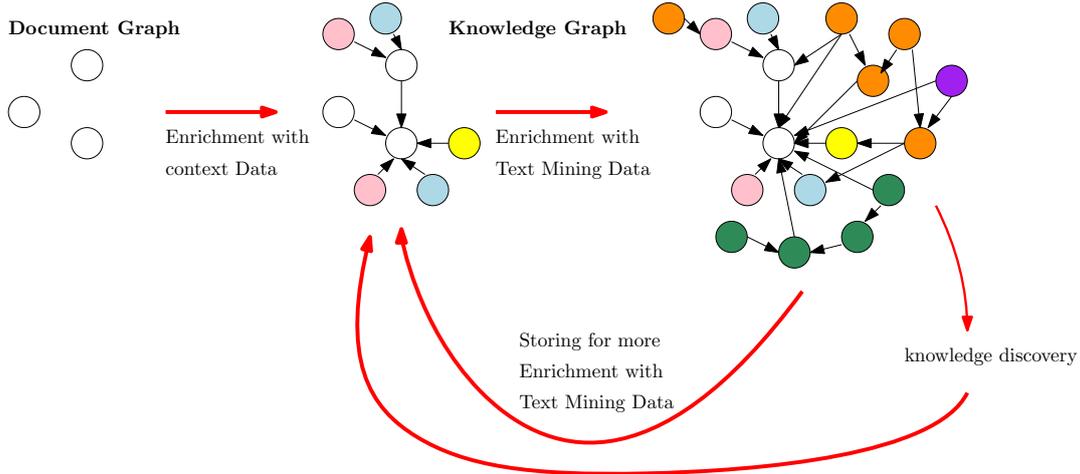


Fig. 1: Proposed workflow to extend a knowledge graph. First starting with a document graph, the basic meta information like authors, keywords etc. are added. This can be used as a basis for text mining which can be used to extend the graph again, for example named entity recognition (NER) may use keywords as a context. Topic detection may also benefit from already assigned keywords, journals or author information. The graph can also be extended by knowledge discovery processes, for example finding parameters of a clinical trial, progression within electronic health records, etc. In any case new context information will be added to the initial graph and improve the input of further algorithms.

the context of knowledge graphs. We demonstrate this novel approach by applying it to common data sources. After that, we will give a detailed list of issues that have to be addressed.

## I. PRELIMINARIES

We define knowledge graphs  $G = (E, R)$  where the set of nodes  $E$  consist of entities  $e \in E$  coming from a formal structure like an ontology  $E_i = (V(E_i), R(E_i))$ .  $E$  is a union of ontologies  $E = \{E_1, \dots, E_n\}$ . The relations  $r \in R$  can be ontology relations, thus in general we can say every ontology  $E_i$  which is part of the data model is a subgraph of  $G$  which means  $E_i \subseteq G$ . In addition, we allow inter-ontology relations between two nodes  $e_1, e_2$  with  $e_1 \in E_1, e_2 \in E_2$  and  $E_1 \neq E_2$ . More general we define  $R = \{R_1, \dots, R_n\}$  as list of either inter-ontology and intra-ontology relations. Both  $E$  as well as  $R$  are finite discrete spaces.

Every entity  $e \in E$  may have some additional meta information which need to be defined with respect to the application of the knowledge graph. For instance there might be several node sets (some ontologies, some document spaces (patents, research data, ...), author sets, journal sets, ...)  $E_1, \dots, E_n$  so that  $E_i \subset E$  and  $E = \cup_{i=1, \dots, n} E_i$ . The same holds for  $R$  where several context relations might come together like "is cited by", "has annotation", "has author", "is published in", etc.

We define a finite, discrete set  $C = \{c_1, \dots, c_m\}$  of contexts  $C_i$ . Every node  $e \in G$  and every edge  $r \in R$  may have one ore more contexts  $c \in C$  denoted by  $con(e)$  or  $con(r)$ . It is also possible to set  $con(e) = \emptyset$ . Thus, we have a mapping  $con : E \cup R \rightarrow \mathcal{P}(C)$  to the power set of  $C$ . If we use a quite general approach towards context, we may set  $C = E$ . Thus, every

inter-ontology relation defines context of two entities, but also the relations within an ontology can be seen as context.

Every node set  $E_i \in \{E_1, \dots, E_n\}$  induces a subgraph  $G[E_i] \subset G$ . With  $G^c[E_i] = G[E_i] \cup N(E_i)$  we denote the extended context subgraph which also contains the neighbours  $N(E_i)$  of each node  $e \in E_i$  in  $G$ , which is the context of that node. For a graph drawing perspective, if  $G^c[E_i]$  defines a proper surface, we can think about a graph embedding of another subgraph  $G^c[E_j]$  on  $G^c[E_i]$ .

We can create the metagraph  $M = (C, R')$  of these contexts. Each context is identified by a node in  $M$ . If there is a connection in  $G$  between two contexts, we add an edge  $(c_1, c_2) \in R'$ . This means if  $\exists(v_1, v_2) \in R : c_1 \in con(v_1), c_2 \in con(v_2) \Rightarrow (c_1, c_2) \in R'$  or  $\exists(v_1, v_2) \in R : c_1 \in con((v_1, v_2)), c_2 \in con(v_2) \Rightarrow (c_1, c_2) \in R'$  or  $\exists(v_1, v_2) \in R : c_1 \in con(v_1), c_2 \in con((v_1, v_2)) \Rightarrow (c_1, c_2) \in R'$ . See figure 2 for an illustration.

Adding edges between the knowledge graph  $G$  or a subgraph  $G' = (E', R') \subseteq G = (E, R)$  and the metagraph  $M$  in  $G \cup M$  will lead to a novel graph. This can be either seen as inverse mapping  $con^{-1}(G')$  or as the hypergraph  $\mathcal{H}(G') = (X, \hat{E})$  given by

$$X = E' \cup G^c[E_i]$$

$$\hat{E} = \{\{e_i, e \forall e \in N(e_i)\} \forall e_i \in X\}$$

This graph can be seen as an extension of the original knowledge graph  $G'$  where contexts connect not only to the initial nodes, but also every two nodes in  $G'$  are connected by a hyperedge if they share the same context. See figure 3 for an illustration.

If  $C = E$ , this will lead to new edges in  $G$  enriching the original graph. This step should be done after every additional extension to the graph  $G$ . Thus we need to update both  $G$  as well as  $M$ .

We will denote this hypergraph  $H$  on a knowledge graph  $G$  and a metagraph  $M$  with  $H_{G|M}$ . We might also add multiple metagraphs  $M_1$  and  $M_2$  which will be denoted by  $H_{G|M_1, M_2}$ .

This graph can be seen as an enrichment of the original knowledge graph  $G$  with contexts. It can be used to answer several research questions and can be utilized to find graph-theoretic formulations of research questions.

If the mapping  $con$  is well defined for the domain set the Graph  $H$  can be generated in polynomial time. Since this is in general not the case, this usually contains a data or text mining task to generate contexts from free texts or knowledge graph entities. With respect to the notation described in [7] this problem  $p$  can be formulated as

$$p = \mathbb{D}|R|f : \mathbb{D} \rightarrow \mathbb{X}|err|\emptyset \quad (1)$$

Here, the domain set  $\mathbb{D}$  is explicitly given by  $\mathbb{D} = G$  or – if additional full-texts  $\hat{D}$  supporting the knowledge Graph  $G$  exist –  $\mathbb{D} = \{G, \hat{D}\}$ . In our case the domain subset  $R = \mathbb{D}$ . In this case we need to find a description function  $f : \mathbb{D} \rightarrow \mathbb{X}$  with a description set  $\mathbb{X} = C$  which holds all contexts. To find relevant contexts we need an error measure  $err : \mathbb{D} \rightarrow [0, 1]$ .

We have to consider several research questions. First of all: What are meta information that can be used to generate a context for a new metagraph? Good candidates are authors, citations, affiliation, journal, and MeSH-terms or rather keywords since they are available in most databases. We also need to discuss text mining results like NER, relationship mining etc. Having more general data like study data, genomics, images, etc. we might also consider side effects; disease labels, population labels (male; female; age; social class; etc.). Here we show a proof of concept for less complex text mining meta data. See figure 1, which describes the process of starting with a simple document graph that can be extended with more context data from text mining. We discuss this in more detail within the next section.

The further research questions address the application of this novel approach for both biomedical research as well as text classification and clustering, NLP and knowledge discovery, also with focus on Artificial Intelligence (AI). How can we use the context metagraph to answer biomedical scientific questions? What can we learn from connections between contexts and how do they look like in the knowledge graph? How can we use efficient graph queries utilizing the context? It may also be useful to filter paths in the knowledge graph according to a given context or to generate novel visualizations. A possible question might be to learn about mechanisms linked to co-morbidities or mechanisms being contextualized by drug information. The meta-graph may also contain information about cause-and-effect relationships in the knowledge graph that are “valid” in a biomedical sense under certain conditions. In addition, a contextualization-based on

demographic information or polypharmacy information. We will discuss several use cases within the last section.

## II. METHOD AND PRACTICAL APPLICATION

The following software was written in Java using Spring Boot (see <http://spring.io/projects/spring-boot>) and Spring Data (see <https://spring.io/projects/spring-data>) and integrated in our SCAIView microservice architecture, see [8]. The database backend is a graph database running Neo4j (see <https://neo4j.com/>).

We will illustrate the following methods with example runs on MedLine and Pubmed data. Both sources are already included in the SCAIView NLP-pipeline. PubMed contains 29 million abstracts from biomedical literature, PMC about 4 million full-text articles.

### A. Creating a document and context graph with basic context extraction

The initial step of creating a document and context graph with basic context extraction needs a basic definition of entity sets  $E_1, \dots, E_n$  and their relations.

The articles and abstracts from PubMed and PMC already come with a lot of contextual data. We may set  $E_{Document}$  as the document set containing nodes, each representing one document. In addition, we may add a set  $E_{Source} = \{\text{PubMed}, \text{PMC}\}$  as the source of a document. Thus, each document can be interpreted as context of a data source.

All meta data are stored in new node sets.  $E_{Author}$  stores the set of authors,  $E_{Affiliation}$  their affiliation which is again context for the authors. Another relevant context is the publisher, in our case  $E_{Journal}$ . PubMed stores several classes, for example Books and Documents, Case Reports, Classical Article, Clinical Study, Clinical Trial, Journal Article, Review etc. We store this in  $E_{PublicationType}$ .

Another important context is  $E_{Annotation}$  storing all kind of annotations like named entities or keywords, which come from the MeSH tree, see [9] and [https://www.nlm.nih.gov/mesh/intro\\_trees.html](https://www.nlm.nih.gov/mesh/intro_trees.html). Thus,  $E_{MeSH} \subset E_{Annotation}$  already comes with a hierarchy and edges  $R_{MeSH}$ . The value of MeSH terms and their hierarchy for knowledge extraction was shown in several recent studies like [10]. We will discuss the value of MeSH as controlled vocabulary within the next section. See figure 4 for an illustration of a single document.

All other relations can be added between the sets  $E_i$ , for example  $R_{isCoAuthor}$ ,  $R_{hasAffiliation}$ , etc. With these information given it is – from an algorithmic point of view – quite easy to add all context relations like  $R_{hasDocument}$ ,  $R_{isAuthor}$ ,  $R_{hasAnnotation}$ ,  $R_{hasCitation}$  etc. Edges must also store additional provenance information. See figure 5 for an illustration.

### B. Extending the knowledge graph using NLP-technologies

The initial knowledge graph can be extended by NLP-technologies.

Terminologies and Ontologies are a widely considered topic in research during the last years. They play an important role in

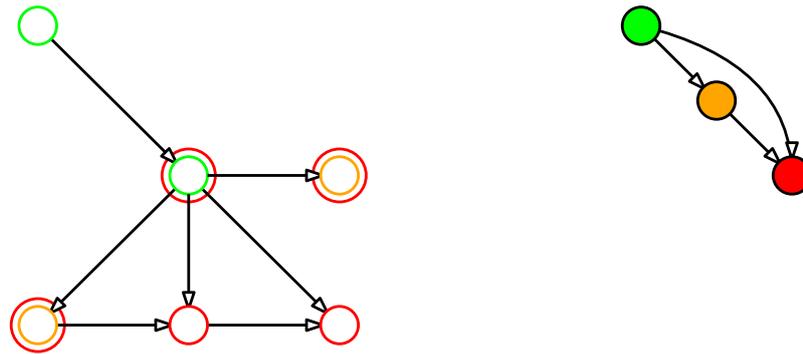


Fig. 2: Illustration of a *knowledge graph with context* (left). The context is illustrated by colors surrounding nodes. At the right the corresponding *context metagraph*. Every context in the knowledge graphs refers to a node in the metagraph and the edges describe if in the original graph a edge from one context to the next exist. Contexts may also be added to edges.

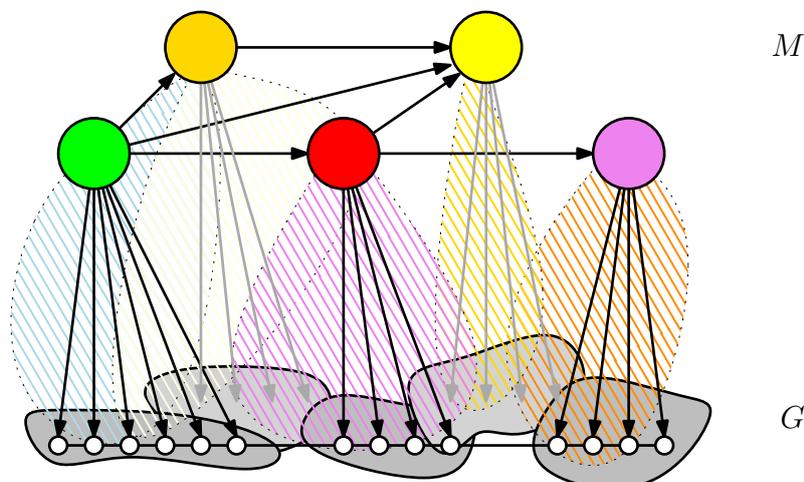


Fig. 3: This figure describes the hypergraph  $\mathcal{H}(G') = (X, \hat{E})$  between the context metagraph  $M$  and the original knowledge graph  $G$  or a subgraph  $G' \subset G$ . This graph is sorted by contexts. The hyperedges, illustrated by sets and indicated by non-hyperedges, connect nodes with context, but also nodes with the same context.

data and text mining as well as knowledge representation in the semantic web. They become more and more important since data provider publish their data in a semantic web formats, namely RDF ([11]) and OWL ([12]), to increase integratability. The term *terminology* refers to the SKOS meta-model [13] which can be summarized as concepts, unit of thoughts which can be identified, labeled with lexical strings, assigned notations (lexical codes), documented with various types of note, linked to other concepts and organized into informal hierarchies and association networks, aggregated, grouped into labeled and/or ordered collections, and mapped to concepts. Several complex models have been proposed in literature and have been implemented in software, see [14]. *Controlled Vocabularies* contain lists of entities which may be completed to a *Synonym Ring* to control synonyms. *Ontologies* also present properties and can establish associative relationships which can also be done by *Thesauri* or *Terminologies*. See [15] and [16] for a complete list of all models.

Here we define Terminologies similar to Thesauri as a set of concepts. They form a *DAG* (Directed Acyclic Graph) with child and parent concepts. In addition, we have an associative relation which identifies similar or somehow related concepts. Each concept has one or more labels. One of them is the preferred identifier, all others are synonyms. To sum up, using ontologies or terminologies for NER, we will have a hierarchy within this ontology. But we may not only consider ontologies and terminologies, but also controlled vocabularies like MeSH. Here we have additional annotations with a different provenance, one coming as keywords with the data, one obtained from NER.

Another example is the Alzheimer's Disease Ontology (ADO, see [17])  $E_{ADO}$  or the Neuro-Image Terminology (NIFT, see [18])  $E_{NIFT}$  coming with their hierarchy  $R_{ADO}$ ,  $R_{NIFT}$ . The process of NER will lead to another context relation  $E_{hasAnnotation}$ . Since not all ontologies or terminologies are described in RDF or OBO format we have to add data from

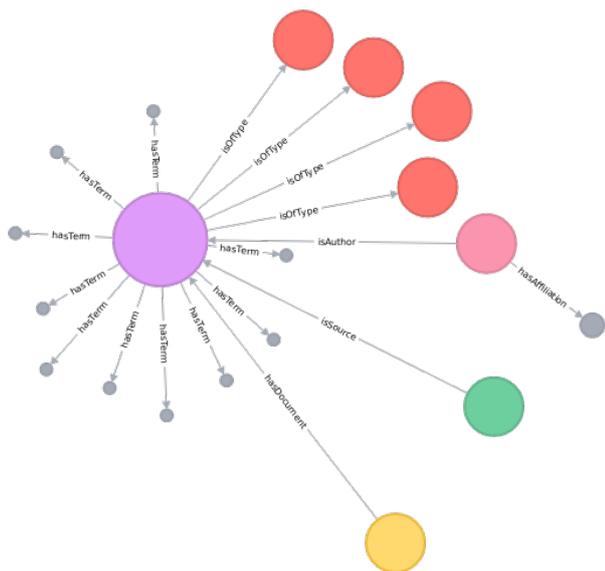


Fig. 4: This figure is an illustration of a single document within the context graph. The document node (purple) has several gray annotation nodes, four red publication type nodes, an orange author node with a gray affiliation. The source (PubMed) is annotated in a green node, the journal in a yellow node.

multiple sources. This is done by a central tool providing all ontology data.

Another context data useful for knowledge extraction are citations, thus edges  $R_{hasCitation}$  between two nodes in  $E_{Document}$ . The data from PMC already stores citation data with unique identifiers (PubMed IDs). Some data is available with WikiData, see [19] and [20]. Other sources are rare, but exist, see [21]. Especially for PubMed a lot of research is working on this difficult topic, see for example [22].

In addition we can consider relational information between entities. For example BEL statements already form knowledge graphs of semantic triples that consist of concepts, functions and relationships [2]. To tackle such complex tasks they constantly gather and accumulate new knowledge by performing experiments, and also studying scientific literature that includes results of further experiments performed by researchers. Existing solutions are mainly based on the methods of biomedical text mining to extract key information from unstructured biomedical text (such as publications, patents, and electronic health records). Several information systems have been introduced to support curators generating these networks. BELIEF, is one workflow generated for this purpose. BELIEF build BEL like statements semi-automatically from retrieving publications from a relevant corpus generation system called SCAIView, see [23] and [24].

Figure 6 illustrates the relations "*Levomilnacipran*" inhibits "*BACE1*", "*BACE1*" improves "*Neuroprotection*" and "*BACE1*" improves "*Memory*" found with relation extraction on named

entities in a document. It is easy to see that context for a document is now also context for the relations and vice versa. If an entity within the relation has synonyms or is found within another document with a different context, this might lead to a deeper knowledge about the statement. Vice versa the context of the document, for example if the knowledge was found within a clinical trial, is a context to the statements.

### III. APPLICATIONS

We will first of all discuss some missing data or data integration problems as well as technical issues which need to be solved. Afterwards we will give an outlook on NLP-based on context information and the impact on answering semantic questions. This is highly related to the FAIRification of research data. This will lead to a short outlook on personalised medicine.

#### A. Missing data

We faced several issues with data integration and missing data. For example some publishers used OCR technologies to convert PDF documents in XML structures. These were usually problematic to process because some fields were missing or wrongly filled.

We have not yet worked on the problem of author and affiliation disambiguation. This is still a widely discussed topic, see [25]. An interesting novel approach – also based on Neo4j database technology – was introduced in [26]. The authors used topological and semantic structures within the graph for author disambiguation. Thus, we plan to integrate state-of-the-art technologies.

In addition performance is a major problem, and the main cause of latency for request. Thus, we had serious problems integrating this framework in our microservice architecture, see [8]. There are several possible explanations for this result, both on technical as well as implementation side. Thus, an important finding was that the storing and retrieval of large knowledge graphs did work. Not surprisingly, for giant and very dense knowledge graphs we need to find another solution. We could either improve the database backend by establishing a polyglot persistence architecture or use existing graph databases like Cray Graph Engine, see [27]. This choice has important implications for the further developing of this architecture, for example SPARQL has more limitations than Cypher. This is an important issue for future research.

However, these results were very encouraging and we will discuss some more topics for further research.

#### B. Context-based NLP

This novel system extends our knowledge and the availability of context data. Context data is a very important foundation for text mining [1]. For example, context-based NER was discussed by [28] and there is still ongoing research, for example on content-aware attributed entity embedding (CAAEE), see [29]. The key strength of our approach is that in every step of text mining and NLP all context data is available and new data will be added. Thus, this system can be used for

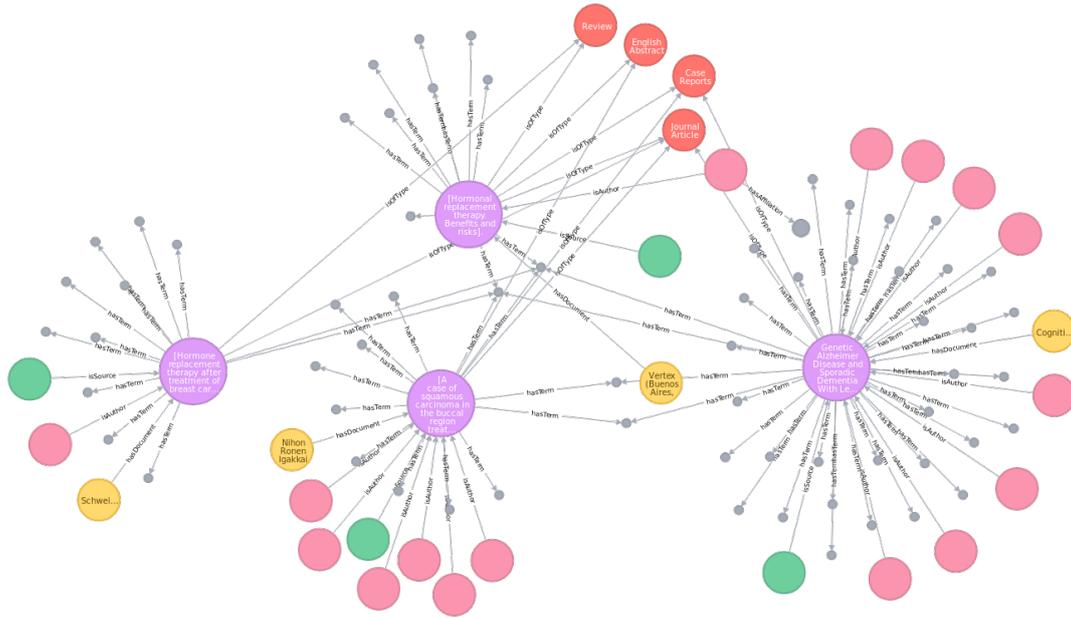


Fig. 5: This figure is an illustration of the initial document and context graph. A PubMed node is the source of document nodes (green). There are several context annotations like article type (red), keywords (gray), authors (orange) and journal (yellow). Authors have additional context (affiliations, gray).

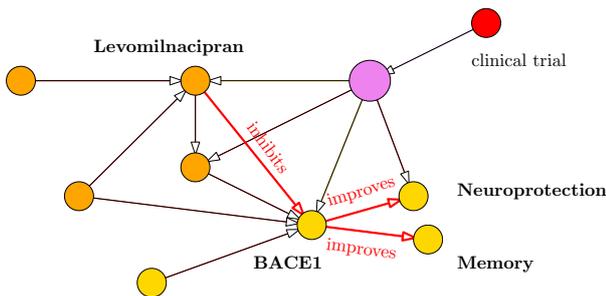


Fig. 6: This figure is an illustration of biological knowledge within the context graph. The document node (purple) has several yellow and orange annotation nodes which come from different terminologies found with NER. The relation extraction task found the relation "Levomilnacipran" inhibits "BACE1", "BACE1" improves "Neuroprotection" and "BACE1" improves "Memory". These relations are illustrated with red edges. Since the document describes a clinical trial, this is also a context for the relations as well.

both building and validating Machine Learning (ML) and AI approaches.

Of course the novel context data is not only suitable for NER, but also for relation extraction. For example [30] proposed a novel approach to context-based relation extraction. Although our example is based on a small data set, the findings suggest that a lot of existing data can be utilized as context data. For example entities annotated by NER or manually curated BEL statements may be applied as context.

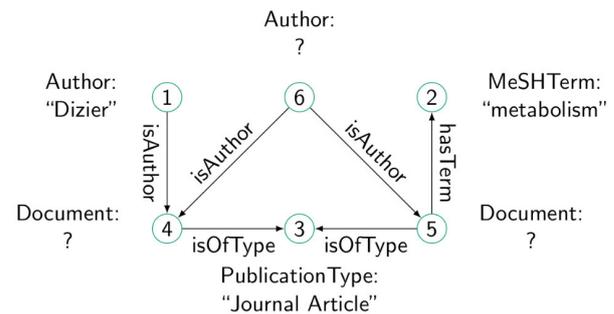


Fig. 7: This figure illustrates a more complex semantic subgraph to query the knowledge graph. We search for two documents having the same author and both of publication type "Journal Article". The first document should have an author called "Dizier", the second one a MeSH Term called "metabolism".

Thus this research has several practical applications. Firstly, it leads to validation and datasets for ML and AI approaches towards text mining. Further work needs to be done to investigate how this data can be used systematically. Secondly, it generalizes the idea of context so that it can be used for semantic questions.

### C. Answering semantic questions and FAIRification of data

Semantic questions can be formulated as subgraph structures of the initial knowledge graphs. For example we may ask: "Which articles have been authored by Pacheco?". This will

lead to a subgraph with two nodes  $v_1, v_2$  where  $v_1 = \text{Pacheco}$  and an edge  $(v_1, v_2) = \text{isAuthor}$ . We may think of much more complex examples, see figure 7 for an example.

In general these semantic subgraph queries (or: graph queries) have an input  $Q = (V, E) \subset G$  and output all subgraphs  $H \subset G$  with  $H \simeq Q$ . Thus, the problem of answering semantic questions is a generalization of the subgraph isomorphism problem. We know already subgraph isomorphism is NP-hard, see [31]. It would be interesting to find a general formulation of the generalization or restrictions that can be applied to this problem. Since Cypher already provides us with the possibility to query graph substructure, further research might explore the runtime or might lead to novel heuristics to solve this efficiently.

Whilst this work did not consider the impact of novel ontologies and terminologies, it did substantiate the impact of them on context data. This is an interesting step towards the FAIRification of data. Wilkinson introduced his FAIR guiding principles in [32] referring to the findability, accessibility, interoperability, and reusability of data, especially for research data. A consequent application of the context idea leads to meta data as context on data which can afterwards be used to make meta data searchable even if the data itself is protected. Thus, the inclusion of context into an information system like SCAIView will make the data findable and accessible. In addition, if interoperable ontologies are available, this data will also be interoperable. This will already solve the three out of four issues addressed by FAIR data.

#### D. Perspectives for Personalised Medicine

Hypothesis generation and knowledge discovery on biomedical data are widely used in medical research and digital health. For example researchers search for genomic or molecular patterns, diagnosis or build longitudinal models. In addition, the massive data available build the basis for a multitude of predictive and personalised medicine ML and AI approaches. A reasonable approach to tackle reproducible research in predictive medicine could be to use a standardized and FAIR context graph for biomedical research data. Thus, it would be necessary to annotate not only biomedical literature but also research data like molecular data, imaging data, genomics and electronic health records (EHR) with context information.

This information system can be used to retrieve data by context (cohort size, settings, results, ...) and by content (imaging data, genomic or molecular measures, ...). For example, this system may answer questions like “Give me a clinical trial to reproduce my results or to apply my model” or “Give me literature for phenotype A, disease B age between C and D and a CT-scan with characteristic E”.

Here we presented a novel approach that annotates research data with context information. The result is a knowledge graph representation of data, the context graph. It contains computable statement representation (e.g. RDF or BEL). This graph allows to compare research data records from different sources as well as the selection of relevant data sets using graph-theoretical algorithms.

## IV. CONCLUSION AND OUTLOOK

Here we discussed a proof-of-concept of a biomedical knowledge graph combining several sources of data as context to each other. We processed data from PubMed and PMC. This initial knowledge graph was extended with results from text mining and NLP-tools already included in our software. Thus, we were able to provide both small datasets as well as large collections of data.

We faced several issues with data integration and missing data, for example because the input data had a bad quality. In addition we have not yet worked on the problem of author and affiliation disambiguation. This directly leads to the question how our approach can be evaluated. For every kind of input data another evaluation method needs to be established. Without this, the quality of the knowledge graph is directly linked to the quality of input data. Before establishing a productive system, this question needs to be properly addressed.

We introduced several applications, for example context-based NLP, answering semantic questions and FAIRification of data, perspectives for Personalised Medicine. The generalisability of these ideas is subject to certain limitations. For instance, the question of interoperable ontologies or ontologies covering the issues of interoperability of data is still not examined. In addition, there is still no FAIR-data information system available.

This has thrown up many questions in need of further investigation. Nevertheless, it is not keen to make an outlook on the impact of such a FAIR and semantic information system and data structure on context data for personalised medicine.

## V. ACKNOWLEDGMENTS

We thank Tim Steinbach for providing some illustrations to this work. In addition, we thank Marc Jacobs and Alexander Esser for carefully revising the manuscript.

## REFERENCES

- [1] C. C. Aggarwal and C. Zhai, “An introduction to text mining,” in *Mining text data*. Springer, 2012, pp. 1–10.
- [2] J. Fluck, A. Klenner, S. Madan, S. Ansari, T. Bobic, J. Hoeng, M. Hofmann-Apitius, and M. Peitsch, “Bel networks derived from qualitative translations of bionlp shared task annotations,” in *Proceedings of the 2013 Workshop on Biomedical Natural Language Processing*, 2013, pp. 80–88.
- [3] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig *et al.*, “Gene ontology: tool for the unification of biology,” *Nature genetics*, vol. 25, no. 1, p. 25, 2000.
- [4] D. S. Wishart, Y. D. Feunang, A. C. Guo, E. J. Lo, A. Marcu, J. R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda *et al.*, “Drugbank 5.0: a major update to the drugbank database for 2018,” *Nucleic acids research*, vol. 46, no. D1, pp. D1074–D1082, 2017.
- [5] K. Khan, E. Benfenati, and K. Roy, “Consensus qsar modeling of toxicity of pharmaceuticals to different aquatic organisms: Ranking and prioritization of the drugbank database compounds,” *Ecotoxicology and environmental safety*, vol. 168, pp. 287–297, 2019.
- [6] C. Haupt, P. Groth, and M. Zimmermann, “Representing text mining results for structured pharmacological queries,” *ISWC*, 2011.
- [7] J. Dörpinghaus, J. Darms, and M. Jacobs, “What was the question? a systematization of information retrieval and nlp problems,” in *2018 Federated Conference on Computer Science and Information Systems (FedCSIS)*. IEEE, 2018.

- [8] J. Dörpinghaus, J. Klein, J. Darms, S. Madan, and M. Jacobs, "Scaiview – a semantic search engine for biomedical research utilizing a microservice architecture," in *Proceedings of the Posters and Demos Track of the 14th International Conference on Semantic Systems - SEMANTiCS2018*, 2018.
- [9] F. B. Rogers, "Medical subject headings," *Bulletin of the Medical Library Association*, vol. 51, pp. 114–116, 1963.
- [10] H. Yang and H. Lee, "Research trend visualization by mesh terms from pubmed," *International journal of environmental research and public health*, vol. 15, no. 6, p. 1113, 2018.
- [11] R. Cyganiak, D. Wood, and M. Lanthaler, "RDF 1.1 concepts and abstract syntax," W3C, W3C Recommendation, Feb. 2014, <http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>.
- [12] P. Patel-Schneider, S. Rudolph, M. Krötzsch, P. Hitzler, and B. Parsia, "OWL 2 web ontology language primer (second edition)," W3C, Tech. Rep., Dec. 2012, <http://www.w3.org/TR/2012/REC-owl2-primer-20121211/>.
- [13] E. Summers and A. Isaac, "SKOS simple knowledge organization system primer," W3C, W3C Note, Aug. 2009, <http://www.w3.org/TR/2009/NOTE-skos-primer-20090818/>.
- [14] M. Zeng, M. Hlava, J. Qin, G. Hodge, and D. Bedford, "Knowledge organization systems (kos) standards," *Proceedings of the Association for Information Science and Technology*, vol. 44, no. 1, pp. 1–3, 2007.
- [15] "Guidelines for the construction, format, and management of monolingual controlled vocabularies," National Information Standards Organization, Baltimore, Maryland, U.S.A., Standard, 2005.
- [16] M. Zeng, "Knowledge organization systems (kos)," vol. 35, pp. 160–182, 01 2008.
- [17] A. Malhotra, E. Younesi, M. Gündel, B. Müller, M. T. Heneka, and M. Hofmann-Apitius, "Ado: A disease ontology representing the domain knowledge specific to alzheimer's disease," *Alzheimer's & Dementia*, vol. 10, no. 2, pp. 238 – 246, 2014.
- [18] A. Iyappan, E. Younesi, A. Redolfi, H. Vrooman, S. Khanna, G. B. Frisoni, and M. Hofmann-Apitius, "Neuroimaging feature terminology: A controlled terminology for the annotation of brain imaging features," *Journal of Alzheimer's Disease*, vol. 59, no. 4, pp. 1153–1169, 2017.
- [19] J. Voß, "Classification of knowledge organization systems with wiki-data," in *NKOS@ TPD L*, 2016, pp. 15–22.
- [20] D. Vrandečić, "Toward an abstract wikipedia," in *31st International Workshop on Description Logics (DL)*, ser. CEUR Workshop Proceedings, M. Ortiz and T. Schneider, Eds., no. 2211, Aachen, 2018. [Online]. Available: <http://ceur-ws.org/Vol-2211/#paper-03>
- [21] A. Obwald, J. Schöpfel, and B. Jacquemin, "Continuing professional education in open access. a french-german survey," *LIBER Quarterly. The journal of the Association of European Research Libraries*, vol. 26, no. 2, pp. 43–66, 2015.
- [22] A. Volanakis and K. Krawczyk, "Sciride finder: a citation-based paradigm in biomedical literature search," *Scientific reports*, vol. 8, no. 1, p. 6193, 2018.
- [23] S. Madan, S. Hodapp, P. Senger, S. Ansari, J. Szostak, J. Hoeng, M. Peitsch, and J. Fluck, "The bel information extraction workflow (belief): evaluation in the biocreative v bel and iat track," *Database*, vol. 2016, 2016.
- [24] S. Madan, J. Szostak, J. Dörpinghaus, J. Hoeng, and J. Fluck, "Overview of bel track: Extraction of complex relationships and their conversion to bel," *Proceedings of the BioCreative VI Workshop*, 2017.
- [25] J. Kim, "Correction to: Evaluating author name disambiguation for digital libraries: a case of dblp," *Scientometrics*, vol. 118, no. 1, pp. 383–383, 2019.
- [26] V. Franzoni, M. Lepri, and A. Milani, "Topological and semantic graph-based author disambiguation on dblp data in neo4j," *arXiv preprint arXiv:1901.08977*, 2019.
- [27] C. D. Rickett, U.-U. Haus, J. Maltby, and K. J. Maschhoff, "Loading and querying a trillion rdf triples with cray graph engine on the cray xc," in *Cray User Group*, 2018.
- [28] D. Nadeau and S. Sekine, "A survey of named entity recognition and classification," *Linguisticae Investigationes*, vol. 30, no. 1, pp. 3–26, 2007.
- [29] D. Cai and G. Wu, "Content-aware attributed entity embedding for synonymous named entity discovery," *Neurocomputing*, vol. 329, pp. 237–247, 2019.
- [30] P. Prajapati and P. Sivakumar, "Context dependency relation extraction using modified evolutionary algorithm based on web mining," in *Emerging Technologies in Data Mining and Information Security*. Springer, 2019, pp. 259–267.
- [31] S. A. Cook, "The complexity of theorem-proving procedures," in *Proceedings of the third annual ACM symposium on Theory of computing*. ACM, 1971, pp. 151–158.
- [32] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne *et al.*, "The fair guiding principles for scientific data management and stewardship," *Scientific data*, vol. 3, 2016.

# Towards semantic-rich word embeddings

Grzegorz Beringer, Mateusz Jabłoński, Piotr Januszewski,  
Andrzej Sobecki, Julian Szymański  
Faculty of Electronic Telecommunications and Informatics  
Gdańsk University of Technology, Gdańsk, Poland  
Email: julian.szymanski@eti.pg.edu.pl

**Abstract**—In recent years, word embeddings have been shown to improve the performance in NLP tasks such as syntactic parsing or sentiment analysis. While useful, they are problematic in representing ambiguous words with multiple meanings, since they keep a single representation for each word in the vocabulary. Constructing separate embeddings for meanings of ambiguous words could be useful for solving the Word Sense Disambiguation (WSD) task.

In this work, we present how a word embeddings average-based method can be used to produce semantic-rich meaning embeddings. We also open-source a WSD dataset that was created for the purpose of evaluating methods presented in this research.

## I. INTRODUCTION

Word embedding methods, that map the vocabulary words to low-dimensional continuous space, have been widely applied to various natural language processing (NLP) problems. They are commonly used as the input representation of words, replacing high-dimensional one-hot encodings, and have been shown to improve the performance in tasks such as syntactic parsing[1] and sentiment analysis[2].

In word embedding methods, such as word2vec[3] or GloVe[4], each word in the vocabulary has exactly one representation. While it is enough for most words, it is problematic for ambiguous words, which can contain more than one meaning. For example, consider the following examples with the word *tree*, extracted from Wikipedia articles:

(a) *The olive, known by the botanical name *Olea europaea*, meaning "European olive", is a species of small **tree** in the family Oleaceae [...]*

(b) *Many theories of syntax and grammar illustrate sentence structure using phrase **trees**, which provide schematics of how the words in a sentence are grouped and relate to each other.*

(c) *Upon completion of listing all files and directories found, **tree** returns the total number of files and directories listed.*

All three sentences mention the word *tree* (or *trees*), but the meaning differs based on context - (a) means tree as a forest plant, (b) tree as a parse tree, (c) tree as a command in Unix systems.

For many applications, such as improving relevance of search engines, anaphora resolution or coherence, identifying which meaning is used, based on context, is important. This task is called Word Sense Disambiguation (WSD) and is an open problem in NLP domain. Word embeddings cannot be applied to WSD out-of-the-box, since they cannot differentiate between multiple meanings of an ambiguous word.

In this work, we propose a method to create semantically rich embeddings for each *keyword* (ambiguous word together with meaning, e.g. *tree (structure)*, *pool (computer science)*), by averaging embeddings of the ambiguous word and words describing its meaning. We evaluate this approach on a WSD task, gathered from Wikipedia articles (III). Finally, we discuss our results and propose future work (V-A).

## II. RELATED WORK

There have been many methods of creating semantically meaningful word representations. Global matrix factorization methods, such as latent semantic analysis (LSA)[5], use matrix factorization to perform dimensionality reduction on a large term-frequency matrix, that captures statistical information about the corpus. As the result, we receive word and document embeddings, which are parametrized by the number of topics we want to extract from the documents, and which can be used to find similarities between different words and documents.

Other approach to creating word embeddings is to take only local context into account, without using global statistics. Example of this is word2vec[3], where a shallow neural network is trained to either predict context words based on the current word (skig-gram), or predict current word based on context words (continuous bag-of-words). Continuous representations of words are then extracted from the hidden layer of the trained network. FastText[6] improves upon skip-gram method, by representing each word as a bag of character n-grams, which provides more flexibility and has an added benefit of the ability to compute word representations for words unseen during training.

Global Vectors (GloVe)[4] combine both global matrix factorization and local context window methods, by training word vectors on co-occurrence matrix, so that their differences predict co-occurrence ratios.

Word embeddings can be also extracted from a trained language model[7]. Recently, methods like ELMo[8] or BERT[9] were shown to achieve great results in many NLP tasks. They produce deep contextualized word embeddings by using internal states of a trained language model pretrained on large corpus of text. Since models used are bidirectional (LSTM for ELMo, Transformer for BERT), the word embedding is conditioned on its left and right context, achieving flexible vector representations that could be used to disambiguate words.

Jacobacci et al.[10] were the first to try to use word embeddings for Word Sense Disambiguation. They consider four different strategies for integrating a pre-trained word embeddings as context representation in a supervised WSD system: concatenation, average, fractional and exponential decay of the vectors of the words surrounding a target word. Peters et al.[11] create word representations that differ from traditional word embeddings in that each token is assigned a representation that is a function of the entire input sentence. They use vectors derived from a bidirectional LSTM that is trained with a coupled language model objective on a large text corpus.

The most usual baseline for WSD task is the Most Frequent Sense[12] (MFS) heuristic, which selects for each target word the most frequent sense in the training data. Recent growth of sequence learning techniques using artificial neural networks contributed to WSD research: Raganato et al.[13] propose a series of end-to-end neural architectures directly tailored to the task, from bidirectional Long Short-Term Memory (LSTM) to encoder-decoder models. Melamud et al.[14] also use bidirectional LSTM in their work. They use large plain text corpora to learn a neural model that embeds entire sentential contexts and target words in the same low-dimensional space, which is optimized to reflect inter-dependencies between targets and their entire sentential context as a whole.

### III. DATASET

For the purpose of constructing semantic-rich word embeddings, we manually gathered usage examples for 6 ambiguous words, 4 to 7 meanings each (28 meanings in total). Ambiguous word together with its meaning constitutes a *keyword*, which we use as a separate class when identifying the closest meaning given some context.

We chose ambiguous words based on the number and variety of meanings it had. Meanings themselves were chosen to cover a range of topics (e.g. *tree (forest)*, *tree (family)*, *trees (folk band)*, *tree (command)*). We also tried to look for meanings that are semantically related and can occur in similar context (and in turn be difficult for the model to differentiate between), e.g. *tree (structure)*, *tree (parse)*, *tree (decision)* or *nails (new wave band)*, *nails (hardcore punk band)*. Lastly, we added some keywords, that we suspected to be really underrepresented in the word embedding of the ambiguous word, e.g. *Mars* as the pop singer Bruno Mars (*mars (bruno singer)*) or *pool* as the computer science term (*pool (computer science)*).

Usage examples for keywords were gathered mostly from Wikipedia, using *What links here* utility, which lists all Wikipedia pages that link to a specific article. We used these links to search for usages of our keywords in context. We found that *What links here* utility has some limitations. Many articles linked to the keyword do not use that keyword in text at all or just list it in "See also" section, which does not provide good context around the keyword for the model to improve on. Moreover, some keywords do not have enough usage examples

that can be found on Wikipedia alone. In such cases, other websites were used to find proper usage examples.

The dataset is split into training and test set, with 5 training and 10 test examples for each keyword. Each example is stored in plain text, with the ambiguous word marked with "\*" on both sides. For simplicity, only one word is marked in each text, even if more ambiguous word usages can be found. In case we wanted to mark another word in the same text, we could just add the same example twice, with different words marked each time.

The correct keyword for each example, together with a path to file and a link, where the original text was taken from, are stored in CSV files: *train.csv* for training set, *test.csv* for test set (columns: path,keyword,link). Keywords themselves, together with links to their Wikipedia articles, are stored in *keywords.csv*.

Dataset, together with the code to execute experiments from this paper, can be found on our GitHub repository<sup>1</sup>.

### IV. OUR METHOD

*Keyword* is a sequence of words that is composed of the ambiguous word and words describing its specific meaning, e.g. *tree (forest)* that represents tree as a plant (ambiguous word: *tree*, meaning: *forest*) and *tree (structure)* which represents tree as a mathematical structure (ambiguous word: *tree*, meaning: *structure*).

To get the embedding of the keyword, we average embeddings of all the words in the keyword:

$$\mathbf{k} = e(w_1, w_2, \dots, w_N) = \frac{1}{N} \sum_{i=1}^N e(w_i) \quad (1)$$

where  $e(\cdot)$  is the embedding function used and  $w_1, w_2, \dots, w_N$  is a sequence of  $N$  words that, in this case, constitutes a keyword.

Example for keyword *tree (forest)*:

$$\mathbf{k}_{tree(forest)} = e(tree, forest) = \frac{e(tree) + e(forest)}{2} \quad (2)$$

*Context* is a sequence of words, extracted from some text, which contains an ambiguous word and words surrounding it in text. It is parametrized by **context length**  $l$ , which specifies how many words from both sides of the ambiguous word are taken into consideration.

*Context embedding*  $c$  is also achieved by taking an average of word embeddings (Equation 1). In this case,  $N = 2l + 1$  and  $w_1, w_2, \dots, w_N$  is the context with ambiguous word inside. For some cases  $N < 2l + 1$ , since the ambiguous word may occur at the beginning or end of text example and full context cannot be collected. In this case, we just average the reduced context.

The approach is to use keyword and context embeddings to find the closest keyword given some context, using cosine distance as a similarity metric.

<sup>1</sup><https://github.com/gberinger/automatic-wiki-links>

Table I  
RESULTS ACHIEVED ON THE TEST SET FOR OUR METHOD. COSINE DISTANCE IS MEASURED BETWEEN THE CORRECT KEYWORD AND CONTEXT EMBEDDINGS.

Metric	Our Model
Top-1 accuracy	67%
Top-2 accuracy	85%
Top-3 accuracy	93%

In other words, given an input text and marked ambiguous word within, we extract the context and compute its embedding  $c$ . The keyword, whose embedding is closest to  $c$  w.r.t. cosine distance, is chosen as the ambiguous word’s meaning.

## V. RESULTS OF THE EXPERIMENTS

In our experiments we evaluate how semantic-rich keyword embeddings, perform for the dataset we collected (III), for the our approach. We use a pretrained embedding model from spaCy - *en\_vectors\_web\_lg*, which contains 300-dimensional word vectors trained on Common Crawl with GloVe<sup>2</sup>.

We compare results on the test set with top- $k$  metrics ( $k \in 1, 2, 3$ ), where we check, if the correct keyword is in closest  $k$  keywords given a specific context describe it. We focus mostly on top-1 accuracy, since we are interested if the word is correctly disambiguated.

Due to the high impact of training data order on test results, we take the average score of 30 runs (each with a random order of training data) for each optimization experiment. We evaluate the performance of the proposed model with different context lengths, to see how it affects top- $k$  accuracies (Fig. 1).

We can see, that the model does relatively well. Top-3 accuracy is about 85-90%, which is probably caused by a low number of meanings for each ambiguous word. Top-1 accuracy for shorter context lengths can go as high as 65% but decreases with longer contexts. As suspected (V-A), this is most likely due to the fact, that the average of many word embeddings may make some contexts similar to each other, therefore making it harder to distinguish between some meanings.

The best result w.r.t. top-1 accuracy was achieved with  $l = 3$ , which is why we choose this context length as a starting point for next experiments.

Performance on the test set can be seen in Table I. All metrics improved due to the optimization process of moving correct keywords closer to (and incorrect keywords away from) contexts found in the training set. High top-2 and top-3 accuracies suggest, that the correct keyword is usually relatively close to the context describing it.

It is important to note, that the performance might worsen, if we expand the keyword vocabulary to large-scale experiments, where we have much more possible keywords than 28.

### A. Discussion and future work

We are aware that our method has some limitations. First of all, it may be impossible to achieve the optimal solution, as we can only optimize keyword embeddings, leaving context

embeddings fixed in the multidimensional space. Therefore, it is possible that contexts for specific keyword overlap on contexts for other keyword.

Secondly, the average context embedding may be ambiguous, with a high possibility of two different context being mapped to a similar point in space, especially for longer context lengths. In future work we plan to experiment with different sequence embedding techniques, that might be better suited for this purpose than a flat average.

Finally, we run experiments for a very small number of ambiguous words and meanings. Our method could have problems with a bigger dataset, since it would be much more difficult to separate different keywords.

Constructing semantic-rich embeddings for ambiguous words, by taking the average of embeddings of the ambiguous word and words describing its meaning, and then comparing it with the average embedding of context words describing given keyword, proved to be a surprisingly good approach for the task of disambiguation on the dataset of 28 keywords we collected (III). Our method achieved 67% top-1, 85% top-2 and 93% top-3 accuracy for context length  $l = 3$ . Longer context lengths were shown to decrease the accuracy, since the average of many word embeddings may result in similar embeddings for different contexts.

Further improvements could be sought by using different keyword and context embedding schemes, e.g. weighted average or by using some sentence embedding method. Optimization method itself could be made more stable by applying decay to alpha and beta parameters and by using a validation set for early stopping. It could also be bound to cosine distance between the keyword and context - the bigger the difference, the bigger the update.

It would also be interesting to see, how the suggested approach for constructing semantic-rich embeddings would perform on a large-scale dataset. Such a dataset could be automatically collected from Wikipedia, using disambiguation pages to find ambiguous words and their meanings, and *What links here* utility, to find usage examples for each keyword.

We assume, that the performance of the our model (and thus the quality of keyword embeddings) can be improved, if we provide examples of contexts that the specific keyword appears in. This can be reached by moving keyword embeddings closer to embeddings of contexts they appear in, so for each training example, the correct keyword embedding is shift by a given factor, in the direction of the context embedding, which describes said keyword. Further optimizations can be done, by moving top- $k$  closest keywords that are incorrect given the same context.

## VI. ACKNOWLEDGEMENTS

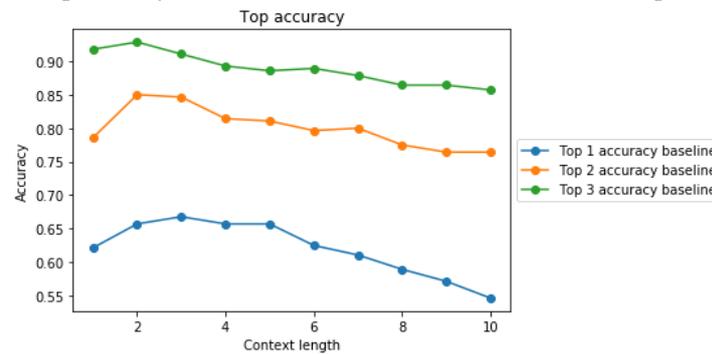
The work has been partially supported by funds of Faculty of Electronics, Telecommunications and Informatics of Gdańsk University of Technology.

## REFERENCES

- [1] R. Socher, J. Bauer, C. D. Manning, and N. Andrew Y., “Parsing with compositional vector grammars,” in *Proceedings of the 51st Annual*

<sup>2</sup>[https://spacy.io/models/en#section-en\\_vectors\\_web\\_lg](https://spacy.io/models/en#section-en_vectors_web_lg)

Figure 1. Top-k accuracies of the our model with different context lengths.



*Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, 2013, pp. 455–465. [Online]. Available: <http://aclweb.org/anthology/P13-1045>

- [2] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Ng, and C. Potts, “Recursive deep models for semantic compositionality over a sentiment treebank,” in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2013, pp. 1631–1642. [Online]. Available: <http://aclweb.org/anthology/D13-1170>
- [3] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” *CoRR*, vol. abs/1301.3781, 2013. [Online]. Available: <http://dblp.uni-trier.de/db/journals/corr/corr1301.html#abs-1301-3781>
- [4] J. Pennington, R. Socher, and C. D. Manning, “Glove: Global vectors for word representation,” in *In EMNLP*, 2014.
- [5] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, “Indexing by latent semantic analysis,” *JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE*, vol. 41, no. 6, pp. 391–407, 1990.
- [6] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, “Enriching word vectors with subword information,” *arXiv preprint arXiv:1607.04606*, 2016.
- [7] Y. Bengio, R. Ducharme, P. Vincent, and C. Janvin, “A neural probabilistic language model,” *J. Mach. Learn. Res.*, vol. 3, pp. 1137–1155, Mar. 2003. [Online]. Available: <http://dl.acm.org/citation.cfm?id=944919.944966>
- [8] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, “Deep contextualized word representations,” feb 2018. [Online]. Available: <http://arxiv.org/abs/1802.05365>
- [9] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” oct 2018. [Online]. Available: <http://arxiv.org/abs/1810.04805>
- [10] I. Iacobacci, M. T. Pilehvar, and R. Navigli, “Embeddings for word sense disambiguation: An evaluation study,” in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, 2016, pp. 897–907. [Online]. Available: <http://aclweb.org/anthology/P16-1085>
- [11] M. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, “Deep contextualized word representations,” in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. Association for Computational Linguistics, 2018, pp. 2227–2237. [Online]. Available: <http://aclweb.org/anthology/N18-1202>
- [12] A. Raganato, J. Camacho-Collados, and R. Navigli, “Word sense disambiguation: A unified evaluation framework and empirical comparison,” in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*. Association for Computational Linguistics, 2017, pp. 99–110. [Online]. Available: <http://aclweb.org/anthology/E17-1010>
- [13] A. Raganato, C. Delli Bovi, and R. Navigli, “Neural sequence learning models for word sense disambiguation,” in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2017, pp. 1156–1167. [Online]. Available: <http://aclweb.org/anthology/D17-1120>
- [14] O. Melamud, J. Goldberger, and I. Dagan, “context2vec: Learning generic context embedding with bidirectional lstm,” in *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*. Association for Computational Linguistics, 2016, pp. 51–61. [Online]. Available: <http://aclweb.org/anthology/K16-1006>

# Languages' Impact on Emotional Classification Methods

Alexander C. Eilertsen, Dennis Højbjerg Rose, Peter Langballe Erichsen,  
Rasmus Engesgaard Christensen, Rudra Pratap Deb Nath  
Department of Computer Science, Aalborg University,  
Selma Lagerlöfs Vej 300, 9220 Aalborg Ø, Denmark  
Email: {aeiler16, drose16, perich16, rech16}@student.aau.dk  
& rudra@cs.aau.dk

**Abstract**—There is currently a lack of research concerning whether Emotional Classification (EC) research on a language is applicable to other languages. If this is the case then we can greatly reduce the amount of research needed for different languages. Therefore, we propose a framework to answer the following null hypothesis: *The change in classification accuracy for Emotional Classification caused by changing a single preprocessor or classifier is independent of the target language within a significance level of  $p = 0.05$ . We test this hypothesis using an English and a Danish data set, and the classification algorithms: Support-Vector Machine, Naive Bayes, and Random Forest. From our statistical test, we got a  $p$ -value of 0.12852 and could therefore not reject our hypothesis. Thus, our hypothesis could still be true. More research is therefore needed within the field of cross-language EC in order to benefit EC for different languages.*

**Keywords:** Sentiment Analysis, Emotional Classification, Text-to-Emotion Analysis, Cross-Language Analysis, Natural Language Processing

## I. INTRODUCTION

The research field of Sentiment Analysis (SA) focuses on textual analysis, concerning the underlying emotions behind language [1]. Emotional information is extracted by using a variety of different methods. This can be used for a number of purposes, e.g. opinion mining during elections.

SA contains the subfield: Emotional Classification (EC), which focuses on classifying the emotions expressed through a medium. For EC, we use the base emotions defined by [2]: joy, trust, fear, surprise, sadness, disgust, anger, and anticipation.

A vast majority of the SA research uses English as the target language. However, it is currently not known whether the results of this research also are applicable to other languages (i.e. cross-language applicability). If the results of SA research based on one target language are applicable to SA for other languages, then that will be very beneficial for SA on non-English languages. We define this area as cross-language EC. To the best of our knowledge no one has conducted research within this area.

Based on this we specify the following null hypothesis: *The change in classification accuracy for Emotional Classification caused by changing a single preprocessor or classifier is independent of the target language within a significance level of  $p = 0.05$ . We test this hypothesis, through an experiment that utilizes a framework we create. This framework consists*

of three overall phases: Preprocessing phase, Classification Phase (CP), and Statical Test Phase (STP). The preprocessing phase consists of three subphases: Common Preprocessing Phase (CPP), Varying Preprocessing Phase (VPP), and Attribute Selection Phase (ASP). This framework serves as a guide for researchers to create experiments with similar structure and purpose as the one we are doing in this study. We do this experiment in order to test whether the effectiveness of different EC methods, trained using tweets, depend on the language being classified.

This experiment uses two data sets; one for Danish and one for English. These data sets consist of posts from the microblogging website *Twitter.com*, called 'tweets'. Tweets are reasonable EC data candidates because they have the purpose of sharing emotions. They are often labeled with keywords, called hashtags, which can be emotional words such as 'happy'. Furthermore, tweets have a character limitation of 280 characters, which entails a higher density of emotions per word.

We compare the differences in impact of changing preprocessors and classifiers on the two data sets, by applying these differences on a two-sided Wilcoxon signed-rank test, from now on referred to as 'Wilcoxon test'.

The result from the Wilcoxon test yields a  $p$ -value of 0.12852, which does not reject our hypothesis. Therefore, it is still possible that EC research on the English language, is applicable to EC on non-English languages. However, since this is only a single experiment, with one non-English language, then more cross-language research is necessary to determine this.

The remainder of the paper is structured in the following way: In Section II we look into previous research within the field of EC. Section III then clarifies the definitions used in this study. Our framework as well as our application of it is defined in Section IV. Details of our experiment are then specified in Section V. In Section VI we present and evaluate our experiment results. The consequences and potential error sources of our results are discussed in Section VII. The conclusion of our study as well as ideas for further research are shown in Section VIII.

## II. RELATED WORKS

During this Section, we introduce a list of EC studies, with a different focus compared to us. We also explain which elements of these studies we use for our experiment.

The main difference between these studies and ours is that while most of these sources examined different preprocessing methods and classification algorithms for the English language, we are comparing preprocessing methods and classification algorithms across multiple languages, in order to check the impact the languages have on their effectiveness.

[3] studied which preprocessing technique yields the highest accuracy using a Naive Bayes Multinomial (NBM) classifier. They used a set of common preprocessors (i.e. preprocessors used in all test cases), and varying preprocessors (i.e. preprocessors which varied whether they were used or not). The combination that yielded the best result, when classifying positive and negative sentences, was the set of common preprocessors and stemming. Using this setup, they were able to achieve an accuracy of 80%.

In [4] they compared accuracies of multiple different  $n$ -gram combinations as well as other features, including preprocessing methods and various lexical resources. Their experiment used LIBLINEAR and NBM as classification algorithms. Based on their research we decided to test the following  $n$ -gram combinations:  $NG = \{1\}$ ,  $NG = \{1, 2\}$ , and  $NG = \{1, 2, 3\}$ .

[5] presented a method for classification using anger, disgust, fear, joy, sadness, and surprise as base emotions, as well as classifying positive, negative, and neutral emotions. The classification was done using a Support-Vector Machine (SVM) classifier with Sequential Minimal Optimization (SMO) calculated on a cluster of computers, and yielded results with accuracies between 65% and 85% depending on the preprocessing methods used. We decided to use some of the preprocessing methods described in [5].

The effectiveness of different SA classification algorithms using tweets was studied by [6]. Based on their research we chose to use Random Forest (RF) and SVM as our classification algorithms. We chose these since we wanted classifiers which performed well and with very different behaviors to cover a wide spectrum of classifiers. RF was overall stable and gave good results, and is chosen as a reliable classifier, whereas SVM showed high performance as a binary classifier, but was shown to be highly data set dependent on 3 class classification.

A framework for detecting emotions in multilingual text was presented by [7]. They developed their emotion extraction system from features that were acquired from different emotional lexicons. Emotions were classified on data gathered from real-time events in different domains, such as sports.

Based on the before mentioned research we chose to use five of the preprocessing methods from [5] and two of the classification algorithms from [6]. We also chose to work with Naive Bayes (NB) as it is a common classification algorithm. We also use the  $n$ -gram preprocessing method with the  $n$ -gram combinations that performed best in [4]:  $NG = \{1\}$ ,  $NG = \{1, 2\}$ , and  $NG = \{1, 2, 3\}$ . While there are many

studies on EC for a single language, there is a lack of research on cross-language EC. The main focus of our research is to address this issue.

## III. PRELIMINARY DEFINITIONS

The definitions we need to clarify are:

- Cross-language: Applying research based on one language to other languages.
- Attribute: Unique word/ $n$ -gram from our data set.
- Instance: A tweet from our data set.
- Class: A base emotion from: {joy, trust, fear, surprise, sadness, disgust, anger, anticipation}[2].
- VPP configuration: A specific combination of preprocessing methods, used in VPP.
- Classification configuration: A combination of a VPP configuration and a classifier.
- Test case configuration: A combination of a classification configuration and a target language.
- Test case: An instance of a test case configuration, including the data set and the results of classifying this data set.

## IV. OUR PROPOSED FRAMEWORK

In this Section, we define the framework for the general point of view as well as how we apply the framework to our experiment.

### A. Framework

The framework is designed to classify a number of test cases. Afterwards, we use a statistical test on these results to evaluate whether the languages used in the data sets have a significant impact on the preprocessors and classifiers being tested.

The input of the framework is a customizable set of data sets in different languages, preprocessing methods, and classification algorithms. Preprocessing methods are divided into common preprocessors and variable preprocessors. Common preprocessors are applied to all test cases, while variable preprocessors are tested as part of the experiment.

We define the framework by three phases: Preprocessing phase, Classification Phase (CP), and Statical Test Phase (STP). The preprocessing phase consists of the following subphases: Common Preprocessing Phase (CPP), Varying Preprocessing Phase (VPP), Attribute Selection Phase (ASP). These phases are visualized in Figure 1. Each test case is going through these phases individually, except STP, which uses the results of the previous phase to evaluate the hypothesis.

The following list provides a general description of each phase, and clarifies its purpose:

- Preprocessing Phase. The purpose of this phase is to make the data sets less complex and faster to classify.
  - Common Preprocessing Phase (CPP). The purpose of this phase is to clean the data set and reduce its size. We do this by removing grammatical elements and combining similar textual elements.
  - Varying Preprocessing Phase (VPP). This phase applies the preprocessing methods that we want to test.

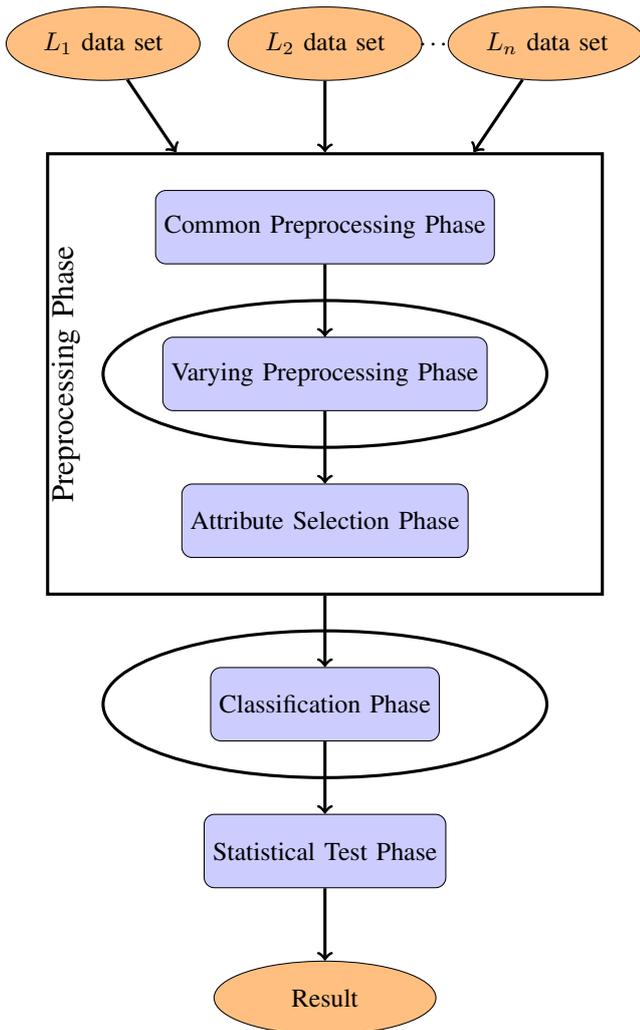


Fig. 1. The process of our framework.  $L_1$  to  $L_n$  represents a minimum of two data sets in different languages to be tested. The black box describes the preprocessing phase which involves the following subphases: CPP, VPP, and ASP. The black ellipses describe the varying parts of our experiment which are changed for each test case.

It consists of multiple preprocessing steps, which are continuously changed based on the VPP configuration of the test case.

- Attribute Selection Phase (ASP). During this phase, we evaluate the preprocessed data set and remove attributes from them in order to reduce classification time.
- Classification Phase (CP). During this phase, we use the preprocessed data set to train and test a classifier.
- Statical Test Phase (STP). During this phase, we use a statistical test on the results gained from CP to evaluate our hypothesis.

### B. Our Application of the Framework

We use a set of Danish tweets and a set of English tweets as the input set for our framework.

Following is a description of the specific methods used for our implementation of each of the phases described in our framework:

1) *Common Preprocessing Phase (CPP)*: Our input for this phase consists of several preprocessing methods which are described below (in execution order):

- **Replace user**  
Replaces a mention of a user, e.g. '@johndoe' with '<user>' in order to unify all references to users.
- **Replace link**  
Replaces a link, e.g. 'pic.twitter.com' with '<link>' in order to unify all references to links, since we do not want to distinguish between links.
- **Remove repeated characters**  
Repeated characters in a word are reduced to a maximum of three repetitions. For example, the word 'happpppppyyy' becomes 'happpyyy'. This is done because a maximum of two adjacent character repetitions can occur naturally, and we assume there is little intensity difference based on the exact amount of repetitions (e.g. 'saaad' and 'saaaaad' have roughly the same intensity). However, we expect a substantial intensity difference between using repeated characters and not which is why up to three repetitions are kept (e.g. 'sad' and 'saaad' have different intensities).
- **Hashtag deletion**  
Hashtags are replaced with the word in the hashtag, e.g. '#sad' becomes 'sad'. Hashtags are often included as words in the text or to summarize the tweet, which is the reason they are kept.
- **Replace emoticons**  
Each emoticon is replaced with an equivalent emoji, thereby reducing the number of attributes. For example, ':D' and ':-D' both become '😊'.
- **Lowercasing**  
All tweets are converted to lowercase.
- **Symbol removal**  
All symbols are removed from the tweets. Commas and semicolons are replaced with <soft>, and additionally <soft> is also added after every string of emojis. Dots, colons, exclamation marks, and question marks are replaced with <hard>. <soft> and <hard> are later used in 'n-gram stop-split' step, described in VPP.

2) *Varying Preprocessing Phase (VPP)*: Our input for this phase consists of the following preprocessing methods (described in execution order):

- **Part-Of-Speech (POS) tagger**  
A POS tagger finds the corresponding word class for each word in the data sets. This is done to focus on typical emotional word classes, i.e. nouns, adjectives, adverbs, and verbs, by removing words from all other classes[8].
- **Stemming**  
Stemming is a process, where each word is converted to its root (e.g. 'walking' becomes 'walk' and 'smiling'

becomes ‘smile’). While some intensity may be lost, the number of attributes are greatly reduced.

- ***n*-gram stop-split**

In this step the *<soft>* and *<hard>* stops are used to split tweets into multiple sets of words, which are split further by *n*-gram before being classified. This means that conjunctions and interposed sentences are taken into account when classifying longer sentences. We use this preprocessor in order to account for the difference in the use of commas between the Danish and the English language. The varying part here is whether *<soft>* is used to split tweets or not while *<hard>* is always used to find splits.

- ***n*-gram**

*n*-gram splits the sets of words acquired in the *n*-gram stop-split preprocessor into smaller sets of words. We test  $NG = \{1\}$ ,  $NG = \{1, 2\}$ , and  $NG = \{1, 2, 3\}$  *n*-gram combination since combinations of multiple *n*-grams received better results than single *n*-grams in [4].

3) *Attribute Selection Phase (ASP)*: Our input for this phase consists of two different methods for removing attributes. Firstly, attributes that only appear in the data set once are removed because they cannot be in the test set and training set at the same time. Besides this we also evaluate the information gain of each attribute, and remove all attributes with an information gain less than 0.00025. This reduced the number of attributes substantially, e.g. for our test case with the most attributes,  $NG = \{1, 2, 3\}$  *English*, we started with 1, 719, 816 attributes, and after running the ASP it had 15, 210 attributes left.

Information gain describes how much information an attribute gives us about the classes. It is calculated using Equation 1, which uses Equation 2 and Equation 3 describing entropy and expected entropy respectively[9].

$$Gain(X) = h(C) - h(C|X) \quad (1)$$

$$h(C) = \sum_{i=1}^n -C_i \cdot \log_2(C_i) \quad (2)$$

$$h(C|X) = \sum_{i=1}^m \frac{|E_i|}{|E|} \cdot h_i(C) \quad (3)$$

In these equations  $C$  is the set of classes  $C = \{C_1, C_2, \dots, C_n\}$ , where  $C_i$  refers to a specific class,  $X$  is an attribute with the domain  $X = \{v_1, v_2, \dots, v_m\}$ , where  $v_i$  refers to a specific value in the domain,  $E_i$  is the set of instances with  $X = v_i$ , and  $h_i(C)$  is the entropy of classes in  $E_i$ .

The domain of our attributes describes how many times the *n*-gram is used in a tweet. However, for the purposes of calculating expected entropy we reduce the domain of all attributes to whether the word is in the tweet or not.

4) *Classification Phase (CP)*: We run all our classifiers using Weka<sup>1</sup>. In order to minimize bias and randomness, we use Weka’s standard parameters, with a 5 fold cross-validation. Which classification algorithm is used depends on the classification configuration from the following options:

- Support-Vector Machine (SVM) - A nonprobabilistic binary classification algorithm. It constructs a hyperplane to separate two classes based on the data points closest to the gap between the classes. We use the SVM optimizer Sequential Minimal Optimization (SMO) for this[10][11].
- Random Forest (RF) - It is also known as random decision forest. RF generates random decision trees which can be used for classification, regression and other purposes[12].
- Naive Bayes (NB) - A simple probabilistic classification algorithm based on applying Bayes’ theorem with strong independence assumptions between the features[13].

5) *Statical Test Phase (STP)*: For our STP, we use a two-sided Wilcoxon signed-rank test[14] on the accuracy difference in pairs of test cases across languages in order to test the following hypothesis:

**Hypothesis:** *The change in classification accuracy for Emotional Classification caused by changing a single preprocessor or classifier is independent of the target language within a significance level of  $p = 0.05$ .*

We cannot use the raw accuracy difference between the languages, since that will only show the difference in difficulty of doing EC on the two languages. Instead we calculate the difference between pairs of classification configurations using our classification results. The difference between the classification configuration pair  $(A, B)$  is calculated as:  $A - B$ . We create a pair of test case configurations  $((A, B)_{Danish}, (A, B)_{English})$  consisting of two pairs of classification configurations.

The test cases representing this test case configuration pair are used as a pair of data points for the Wilcoxon test to make our cross-language comparison. We do this for each pair of classification configurations  $(A, B)$  which only have one difference between them (one varying preprocessor or a different classifier), making up a total of 180 pairs of data points for the Wilcoxon test. These pairs of data points can be seen in Table IV.

We are not using pairs of classification configurations with more than one difference between them since they are already represented through multiple pairs of classification configurations with only one difference;  $(A - C) = (A - B) + (B - C)$ .

## V. EXPERIMENT

During this Section, we specify some details of our experiment, specifically our data extraction process and VPP configurations. We conduct this experiment in order to determine whether the language being classified has impact on the accuracy of EC for a given classification configuration or not.

<sup>1</sup><https://www.cs.waikato.ac.nz/~ml/weka/>

TABLE I

VPP CONFIGURATIONS USED IN OUR EXPERIMENT. CONFIGURATIONS WITH  $NG = \{1\}$  AND NGSS ARE REMOVED AS N-GRAM STOP-SPLIT HAS NO IMPACT ON 1-GRAMS.

Configuration #	Configuration Setup
C1	$NG = \{1\}$
C2	$NG = \{1, 2\}$
C3	$NG = \{1, 2, 3\}$
C4	$NG = \{1\}, ST$
C5	$NG = \{1, 2\}, ST$
C6	$NG = \{1, 2, 3\}, ST$
C7	$NG = \{1\}, POS$
C8	$NG = \{1, 2\}, POS$
C9	$NG = \{1, 2, 3\}, POS$
C10	$NG = \{1, 2\}, NGSS$
C11	$NG = \{1, 2, 3\}, NGSS$
C12	$NG = \{1\}, ST, POS$
C13	$NG = \{1, 2\}, ST, POS$
C14	$NG = \{1, 2, 3\}, ST, POS$
C15	$NG = \{1, 2\}, POS, NGSS$
C16	$NG = \{1, 2, 3\}, POS, NGSS$
C17	$NG = \{1, 2\}, ST, NGSS$
C18	$NG = \{1, 2, 3\}, ST, NGSS$
C19	$NG = \{1, 2\}, ST, POS, NGSS$
C20	$NG = \{1, 2, 3\}, ST, POS, NGSS$

#### A. VPP Configurations

All possible VPP configurations for our experiment are shown in Table I. We use these configurations both for the Danish and the English data set, and for each classifier. This table uses the following abbreviations for describing the types of VPP methods included in each VPP configuration:

- Stemming = ST
- POS tagger = POS
- $n$ -gram = NG
- $n$ -gram stop-split = NGSS

#### B. Data Extraction

For each base emotion, we manually choose hashtags based on synonyms and similar words from these websites<sup>2,3,4</sup>. Then we manually filter the hashtags, based on whether the tweets using the hashtag show the correct emotion. Examples of these hashtags are shown in Table II. We then download the tweets, which include the remaining hashtags, using the python library 'Twint'<sup>5</sup>.

It is important that the data set for each language are as similar as possible. This is to ensure that any difference we detect in the performance of methods is due to linguistic differences rather than other differences in the data sets. In particular, we want the data sets to have equal size and distribution between classes. The English data set is created based on the size of the Danish data set since there are fewer Danish tweets compared to English tweets. For each English

hashtag, we collected a number of tweets equal to  $\frac{1}{10}$  of the number of Danish tweets for the class which the hashtag belongs to. Then from each class of English tweets a number of random unique tweets, equal to the size of the same class of Danish tweets, are selected. This makes the data sets equal in number of tweets for each class, as well as in the total number of tweets.

## VI. EVALUATION

In this Section, we show and discuss the results from our experiment's CP and STP through trends and phenomena that occur.

#### A. Classification Evaluation

For each test case configuration, we calculate accuracy, precision, recall, and F-measure using Weka. Accuracy is a general measure of the quality of the classification. Precision and recall are both measures of relevance, where precision describes how many retrieved items are relevant, and recall describes how many relevant items are retrieved. The values listed are the average precision and recall of the classes. F-measure is the harmonic mean of precision and recall. The values listed are the average F-measure of the classes. Weka calculates these statistics using the following formulas:

$$Accuracy = \frac{|correct\ results|}{|correct\ results \cup incorrect\ results|} \quad (4)$$

$$Precision = \frac{1}{n} \sum_{i=1}^n \frac{|TP(C_i)|}{|TP(C_i)| + |FP(C_i)|} \quad (5)$$

$$Recall = \frac{1}{n} \sum_{i=1}^n \frac{|TP(C_i)|}{|TP(C_i)| + |FN(C_i)|} \quad (6)$$

$$F - measure = 2 \cdot \frac{Precision + Recall}{Precision \cdot Recall} \quad (7)$$

In the above equations,  $C$  is the set of classes  $C = \{C_1, C_2, \dots, C_n\}$ , where  $C_i$  refers to a specific class of emotions, and  $n$  is the number of classes (eight emotions in our case). *correct results* is the set of all results which are classified as the correct class, while *incorrect results* is the set of all results classified as the wrong class.  $TP(C_i)$ ,  $TN(C_i)$ ,  $FP(C_i)$ , and  $FN(C_i)$  describe the set of: true positive-, true negative-, false positive-, and false negative results respectively, for the class  $C_i$ .

We present the results of the CP in Table III. A row in Table III describes which VPP configuration is used, while the columns describe whether accuracy, F-measure, precision, or recall is shown, and which language and classification algorithm is used.

When we observe the results, the following trends appear:

- The average accuracy of the English data set is lower than the average accuracy of the Danish data set. This might be due to the the higher diversity in English tweets, created by the difference in numbers of hashtags, and that the English tweets are written by many different cultures,

<sup>2</sup><https://ordnet.dk/>

<sup>3</sup><https://www.thesaurus.com/>

<sup>4</sup><https://sproget.dk/>

<sup>5</sup><https://github.com/twintproject/twint>

TABLE II  
EXAMPLES AND NUMBER OF HASHTAGS AND TWEETS.

Emotion	# of Hashtags	Danish Hashtag Examples	# of Tweets	Danish Tweet Example
Joy	25	#glad #glæde #fryd	16541	Hold nu op hvor jeg elsker faneblade i Finder i OSX. Det er SÅ genialt! #glæde
Trust	16	#tillid #tillidsfuld #tiltro	4125	Når en fyr viser han er til at stole på #tillid
Fear	35	#frygt #angst #bange	4941	Nu synes jeg godt snart det må falde lidt til ro i Japan tak! #Bekymret
Surprise	26	#overrasket #forundret #forbavset	2224	Så har man set det med.. Unge tabere der leger med lasere... #chokeret
Sadness	33	#ked #kedafdet #deprimeret	20537	Øv, hvor kan man nogen gange blive lidt trist til mode, over de mindste ting #trist
Disgust	39	#beskidt #snavset #gyselig	3889	Er et skridt tættere på at være voksen efter jeg har renset afløb i mit badeværelse! #ulækkert
Anger	34	#vred #arrig #hidsig	4056	Jeg håber at der er en der saver Suarez fuldstændig midt over. #bitter
Anticipation	20	#spændende #nysgerrig #fristende	8859	Jeg fucking håber Lady Gaga kommer til Danmark! #håb

Emotion	# of Hashtags	English Hashtag Examples	# of Tweets	English Tweet Example
Joy	39	#joy #happy #happiness	16541	Final week of semester! #contented
Trust	15	#trust #trustful #admiration	4125	Don't #depend on others when you can #doityourself !
Fear	49	#fear #terror #fright	4941	one of the #worst features about #worrying is that it destroys our ability to #concentrate.
Surprise	25	#surprise #surprising #amazement	2224	@netflix love death & robots is amazing, loving it #astounding
Sadness	47	#grief #sadness #sorrow	20537	I am not sure I care anymore #painful
Disgust	69	#ew #unclean #jealous	3889	I went back to high school for two hours and that's time I can never get back #resent #regret
Anger	43	#angry #anger #mad	4056	I hate Iowa #displeased
Anticipation	28	#anticipation #watchful #expecting	8859	Save the date! Nov 9th to 16th! #expectation

while the Danish tweets primarily are written by Danish people.

- SVM has the highest accuracy, F-measure, precision, and recall, out of all classifiers and across both languages.
- The  $n$ -gram stop-split preprocessor does not make a large difference in the results. There are only a few cases with a noticeable difference, e.g. between C16 and C9, which is  $NG = \{1, 2, 3\}$  POS, with and without NGSS respectively. This might be because most of the  $n$ -grams this preprocessor removes would otherwise have been removed during the ASP.
- The differences in classification effectiveness between  $NG = \{1\}$  and  $NG = \{1, 2, 3\}$  is the opposite of what we expected. The effectiveness of  $NG = \{1\}$  is often higher than the other  $n$ -gram variations for both Danish and English. This suggests that the context gained from adding orders of words is less significant than the noise created by adding more  $n$ -gram attributes.

### B. Statistical Test Evaluation

We compare the classification accuracies, from Table III by applying them on a Wilcoxon test. The basis of this analysis is described in Section IV-B5.

Figure 2 shows the pairs of test case configurations where Table IV shows the setup of each test case configuration i.e. the variables on the x-axis of Figure 2.

In Table IV, test case configuration differences written on the form 'VPP configuration-classifier-classifier' describe two test case configurations with the same VPP configuration but different classifiers. However, test case configuration differences on the form 'VPP configuration-VPP configuration-classifier' describe two test case configurations with one difference in

their VPP configuration but using the same classifier. The corresponding VPP configurations are shown in Table I.

In Figure 2, each point represents the difference between two test cases' accuracy ( $A_{accuracy}, B_{accuracy}$ ), where  $A$  and  $B$  has only one difference between their classification configurations. If a point is positive, then test case  $A$  has a higher accuracy than  $B$ ; if a point is negative, then test case  $A$  has a lower accuracy than  $B$ ; and if a point is 0, then there is no difference between their accuracies.

Each line in Figure 2 represents the accuracy difference between a pair of test case configurations. The red and orange lines represent the English data set, while the blue and cyan lines represent the Danish data set. The special cases where one point is above 0 and the other is below, represent test case configuration pairs where there is a positive accuracy change for one language and a negative change for the other. Orange and cyan represent these special cases. These cases support the rejection of our hypothesis.

Running the Wilcoxon test on our test case configuration pairs results in a  $p$ -value of 0.12852. Our hypothesis is therefore not rejected within a significance level of 0.05. Thus, which classification configuration that performs best might be independent of the languages being classified.

The box plot in Figure 3 shows the variance of the accuracy difference in the data used for the Wilcoxon test. We can see that the English data set has a higher variance, meaning it is more sensitive towards configuration changes. Despite this, both data sets have a median close to 0, which could explain why we cannot reject our hypothesis.

By studying Figure 2, we learn that the biggest differences in accuracy comes from the change of classifier to/from NB. Furthermore, POS tagging on the English data set makes almost

TABLE III

TEST CASE RESULTS: BOLD VALUES ARE THE HIGHEST VALUES WITHIN THE CLASSIFIER AND LANGUAGE COMBINATION WHILE UNDERLINED VALUES ARE THE HIGHEST VALUES WITHIN THE LANGUAGE.

Config.	Accuracy						F-measure					
	Danish			English			Danish			English		
	SVM	NB	RF									
C1	94.66	76.45	91.63	<u>94.24</u>	55.58	87.14	98.33	79.09	96.94	<b>95.54</b>	58.68	93.14
C2	94.45	75.40	89.80	<u>93.97</u>	54.39	83.20	98.32	78.33	96.36	95.23	57.60	90.99
C3	94.40	75.55	88.98	93.88	54.35	81.69	98.30	78.55	96.01	95.19	57.56	89.92
C4	93.97	73.84	90.04	92.81	<b>57.36</b>	83.61	97.89	77.85	96.11	94.38	<b>60.97</b>	90.30
C5	93.68	72.93	87.85	92.48	56.09	80.57	97.94	77.33	95.27	94.11	59.84	88.74
C6	93.66	73.04	86.95	92.46	55.94	79.30	97.95	77.33	94.86	94.12	59.67	87.92
C7	<b>94.71</b>	78.59	<b>92.95</b>	68.46	51.10	67.56	<b>98.35</b>	81.16	<b>97.52</b>	69.26	52.00	68.63
C8	94.54	77.98	92.22	69.01	50.95	67.85	98.33	81.02	97.33	69.86	52.51	69.17
C9	94.53	77.93	91.77	68.98	51.00	67.73	98.31	81.07	97.25	69.66	52.54	69.28
C10	94.45	75.29	89.56	94.01	54.60	83.52	98.34	78.17	96.23	95.26	57.83	91.00
C11	94.40	<b>75.23</b>	88.90	93.93	54.78	<b>91.73</b>	98.31	78.21	95.90	95.26	57.98	<u>97.12</u>
C12	93.40	<b>79.71</b>	91.65	67.55	54.56	67.42	97.76	<b>83.31</b>	97.14	68.38	54.59	68.56
C13	93.20	79.01	90.59	68.48	54.29	67.68	97.89	82.99	96.94	69.23	54.91	69.14
C14	93.14	78.92	90.23	68.44	54.11	67.38	97.89	82.99	96.83	69.14	54.83	68.85
C15	94.55	77.98	92.18	69.03	51.04	67.88	98.32	81.02	97.27	69.68	52.34	69.24
C16	94.52	77.91	81.88	69.01	51.02	67.81	98.32	81.08	90.03	69.62	52.43	69.31
C17	93.71	72.83	87.53	92.55	56.46	80.78	97.93	76.93	95.22	94.19	60.23	89.02
C18	93.66	72.83	86.86	92.51	56.63	79.63	97.94	76.98	94.82	94.12	60.42	88.21
C19	93.22	79.01	90.51	68.33	54.38	67.57	97.88	82.98	96.89	68.99	54.84	68.80
C20	93.17	78.92	90.22	68.45	54.14	67.54	97.88	82.98	96.72	69.08	54.72	68.94
Avg.	<u>94.00</u>	76.46	89.62	<u>80.92</u>	54.14	75.38	<u>98.11</u>	79.97	96.08	<u>82.01</u>	56.32	79.81
Config.	Precision						Recall					
	Danish			English			Danish			English		
	SVM	NB	RF									
C1	98.22	74.80	95.98	<b>95.36</b>	50.65	90.21	98.44	83.94	97.91	95.72	69.77	96.25
C2	98.35	73.98	95.24	94.68	49.20	87.61	98.30	83.24	97.51	95.78	69.47	94.65
C3	98.33	74.39	94.69	94.62	49.02	86.31	98.27	83.23	97.37	95.75	69.73	93.84
C4	97.95	74.56	94.91	94.82	<b>54.39</b>	87.14	97.82	81.46	97.34	93.96	69.45	93.70
C5	98.04	74.10	93.66	94.05	52.54	84.93	97.84	80.89	96.93	94.18	69.55	92.91
C6	98.02	74.09	92.95	94.03	52.11	84.01	97.87	80.88	96.86	94.21	<b>69.84</b>	92.21
C7	98.25	77.35	<b>96.79</b>	61.76	42.78	63.03	<b>98.46</b>	85.39	<b>98.27</b>	78.84	66.28	75.32
C8	98.30	77.73	96.70	62.01	43.67	63.68	98.36	84.62	97.96	80.00	65.85	75.72
C9	98.29	77.89	96.54	61.81	43.66	63.73	98.33	84.53	97.98	79.79	65.99	75.91
C10	<b>98.36</b>	73.71	94.99	94.68	49.57	87.71	98.32	83.25	97.51	95.84	69.41	94.56
C11	98.37	73.79	94.55	94.66	49.76	<b>96.32</b>	98.26	83.21	97.29	<b>95.86</b>	69.46	<b>97.94</b>
C12	97.48	80.57	96.30	62.31	45.71	63.80	98.05	<b>86.27</b>	98.00	75.82	67.78	74.10
C13	97.74	80.65	96.15	62.09	46.49	64.38	98.04	85.50	97.75	78.26	67.08	74.68
C14	97.78	<b>80.71</b>	95.96	61.85	46.39	64.06	97.99	85.41	97.72	78.38	67.06	74.41
C15	98.28	77.73	96.49	61.87	43.11	63.93	98.36	84.62	98.07	79.77	66.63	75.50
C16	98.31	77.88	86.61	61.80	43.08	63.95	98.34	84.56	93.74	79.71	66.97	75.66
C17	98.00	73.39	93.52	94.08	53.31	85.61	97.86	80.86	96.99	94.30	69.26	92.72
C18	97.99	73.53	92.87	94.00	53.55	84.67	97.90	80.79	96.86	94.24	69.36	92.06
C19	97.72	80.63	95.98	61.92	46.06	64.02	98.05	85.49	97.81	77.91	67.79	74.36
C20	97.73	80.68	95.64	61.87	45.84	64.03	98.04	85.44	97.83	78.20	67.89	74.68
Avg.	<u>98.07</u>	76.61	94.82	<u>78.21</u>	48.04	75.66	<u>98.14</u>	83.68	97.39	<u>86.83</u>	68.23	84.56

as large a negative change in accuracy difference as changing classifier to NB. The Danish data set however improves slightly when POS tagging is applied. This effect can be seen in the difference between C1 to C7, C2 to C8, and C3 to C9 for all classifiers.

## VII. DISCUSSION

In this Section, we discuss the consequences of the observations in Section VI-B. First we look at the results of the Wilcoxon test, followed by the effects of classifiers, and language specific tools.

As described in Section VI-B, our Wilcoxon test did not yield any significant results. This suggests that classification configurations react similarly to the Danish and the English data set. However, further research is needed to establish the statement “EC research based on one language is applicable to other languages”.

However, there is a significant difference when NB is applied as a classifier. Using NB, the accuracies of the English data set are between 50% – 58% while the Danish data set’s accuracies are between 72% – 80%. This suggests that there is a relevant difference in EC between the two languages.

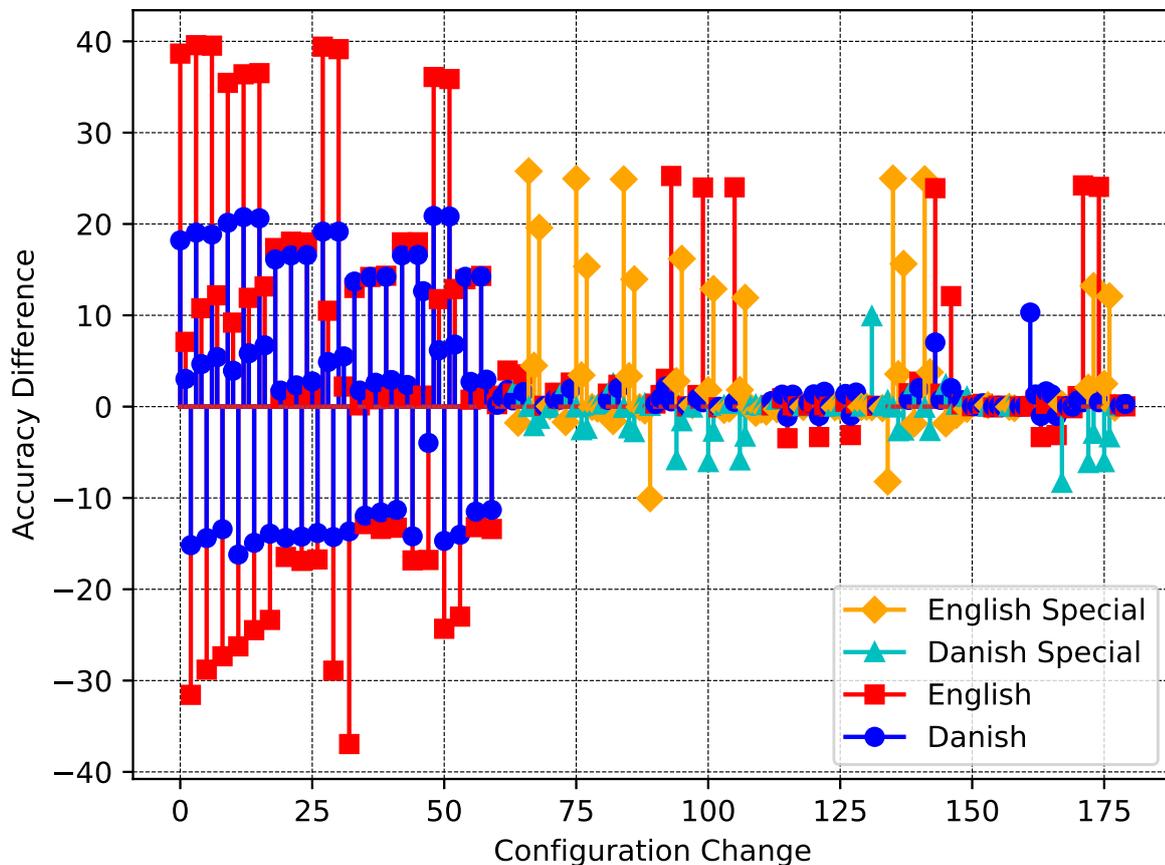


Fig. 2. Data points used in the Wilcoxon test. Configurations can be seen in Table IV. Each data point represents the percentage difference in accuracy between a pair of classification configurations. Data points marked with a blue circle represents the Danish data set and points marked with a red square represents the English data set. The data points with a orange diamonds and cyan triangles represent special cases for Danish and English respectively. These special cases describe where the configuration change had a positive impact on the one language but not with the other.

Another interesting observation we found in Section VI-B is that POS tagging has opposite effects on the two languages. Adding POS tagging made a difference in accuracy between  $-0.58\%$  and  $7.02\%$  on the Danish data set and between  $-1.82\%$  and  $-25.78\%$  on the English data set. The variance is not only higher for the English data set, as shown in Figure 3, the difference is also mostly positive for Danish and always negative for English. This means that the Danish data set benefits from POS tagging while the English data set suffers greatly from it. This suggests that while a lot of the elements of EC are not language dependent, the use of tools designed for a specific language might be language dependent. Therefore, more language specific research in these tools would be beneficial.

#### A. Possible Error Sources

By analyzing our experiment, we find some possible error sources which may have impact on our results.

- There exists non-Danish tweets in the Danish data set since Twitter's language filter is not perfect.
- English tweets are posted more often than Danish tweets, and we download the tweets in chronological descending order of posting time. In order to have the same amount of tweets in the data sets, the Danish data set ends up with a much higher time variance between posts than the English data set. Therefore, the Danish data set probably has a higher variance in how the language is used.
- The hashtags used for gathering tweets have been chosen manually and therefore do not cover all emotional words related to the base emotions.
- There may be differences in how the chosen hashtags are related to the base emotion they are labeled with. There are also 87 more English hashtags than Danish hashtags. This might cause the English data set to be more diverse and therefore possibly harder to classify.
- There are some differences between the Danish and

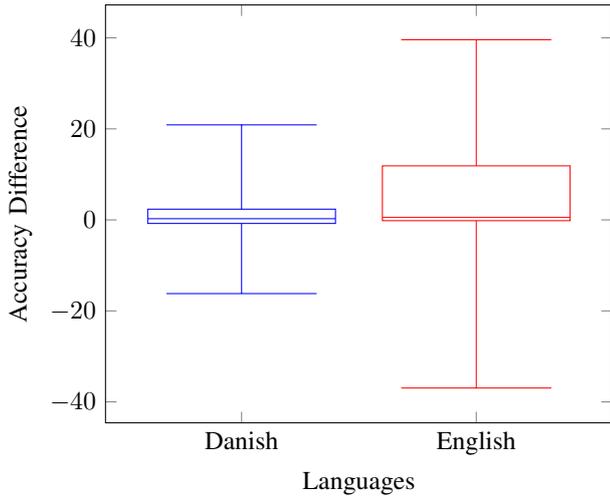


Fig. 3. Boxplot of data in Figure 2 which shows the accuracy difference using the quartiles; {Minimum, Lower Quartile, Median, Upper Quartile, Maximum} for the Danish and English data sets.

English POS tagging and stemming preprocessing methods used in VPP.

- The Danish POS tagger labels nonwords as nouns and the English POS tagger labels nonwords as proper nouns. In the VPP the POS tagging preprocessor keeps nouns but not proper nouns as part of the attributes.

## VIII. CONCLUSION

We have conducted this study in order to test whether the new research field: *cross-language EC* has the potential for reducing the amount of research needed for non-English languages within the field of EC. In Section IV-A, we constructed a framework for testing the classification accuracy of a number of test cases. In Section IV-B, the framework was used to setup our experiment, for the purpose of evaluating our null hypothesis: *The change in classification accuracy for Emotional Classification caused by changing a single preprocessor or classifier is independent of the target language within a significance level of  $p = 0.05$ .* We made this hypothesis in order to answer the more general question: *Do target languages have impact on the effectiveness of EC methods?* Our two-sided Wilcoxon signed-rank test gave a  $p$ -value of 0.12852, and therefore did not reject the hypothesis using data sets constructed from Danish and English tweets. It should be noted that our results are based only on two germanic languages with the common domain Twitter, and thus only covers a small part of the research within cross-language EC. During our experiment, SVM has consistently yielded the best results in contrast to the experiment made by [6], where SVM did not yield consistent results on nonbinary classification. In Section VI, we observed a few interesting characteristics of our results, e.g. POS tagging works well for the Danish data set but not for the English data set. These findings suggest that further research is needed for cross-language EC. We believe our study

TABLE IV

THIS TABLE DESCRIBES THE X-AXIS IN FIGURE 2. EACH X-VALUE DESCRIBES A PAIR OF TEST CASE CONFIGURATIONS WITH ONLY ONE DIFFERENCE.

x	Config. Diff.	x	Config. Diff.	x	Config. Diff.	x	Config. Diff.	x	Config. Diff.	x	Config. Diff.
1	C1-SVM-NB	31	C11-SVM-NB	61	C1-C2-SVM	91	C4-C5-SVM	121	C8-C13-SVM	151	C13-C14-SVM
2	C1-SVM-RF	32	C11-SVM-RF	62	C1-C2-NB	92	C4-C5-NB	122	C8-C13-NB	152	C13-C14-NB
3	C1-NB-RF	33	C11-NB-RF	63	C1-C2-RF	93	C4-C5-RF	123	C8-C13-RF	153	C13-C14-RF
4	C2-SVM-NB	34	C12-SVM-NB	64	C1-C4-SVM	94	C4-C12-SVM	124	C8-C15-SVM	154	C13-C19-SVM
5	C2-SVM-RF	35	C12-SVM-RF	65	C1-C4-NB	95	C4-C12-NB	125	C8-C15-NB	155	C13-C19-NB
6	C2-NB-RF	36	C12-NB-RF	66	C1-C4-RF	96	C4-C12-RF	126	C8-C15-RF	156	C13-C19-RF
7	C3-SVM-NB	37	C13-SVM-NB	67	C1-C7-SVM	97	C5-C6-SVM	127	C9-C14-SVM	157	C14-C20-SVM
8	C3-SVM-RF	38	C13-SVM-RF	68	C1-C7-NB	98	C5-C6-NB	128	C9-C14-NB	158	C14-C20-NB
9	C3-NB-RF	39	C13-NB-RF	69	C1-C7-RF	99	C5-C6-RF	129	C9-C14-RF	159	C14-C20-RF
10	C4-SVM-NB	40	C14-SVM-NB	70	C2-C3-SVM	100	C5-C13-SVM	130	C9-C16-SVM	160	C15-C16-SVM
11	C4-SVM-RF	41	C14-SVM-RF	71	C2-C3-NB	101	C5-C13-NB	131	C9-C16-NB	161	C15-C16-NB
12	C4-NB-RF	42	C14-NB-RF	72	C2-C3-RF	102	C5-C13-RF	132	C9-C16-RF	162	C15-C16-RF
13	C5-SVM-NB	43	C15-SVM-NB	73	C2-C5-SVM	103	C5-C17-SVM	133	C10-C11-SVM	163	C15-C19-SVM
14	C5-SVM-RF	44	C15-SVM-RF	74	C2-C5-NB	104	C5-C17-NB	134	C10-C11-NB	164	C15-C19-NB
15	C5-NB-RF	45	C15-NB-RF	75	C2-C5-RF	105	C5-C17-RF	135	C10-C11-RF	165	C15-C19-RF
16	C6-SVM-NB	46	C16-SVM-NB	76	C2-C8-SVM	106	C6-C14-SVM	136	C10-C15-SVM	166	C16-C20-SVM
17	C6-SVM-RF	47	C16-SVM-RF	77	C2-C8-NB	107	C6-C14-NB	137	C10-C15-NB	167	C16-C20-NB
18	C6-NB-RF	48	C16-NB-RF	78	C2-C8-RF	108	C6-C14-RF	138	C10-C15-RF	168	C16-C20-RF
19	C7-SVM-NB	49	C17-SVM-NB	79	C2-C10-SVM	109	C6-C18-SVM	139	C10-C17-SVM	169	C17-C18-SVM
20	C7-SVM-RF	50	C17-SVM-RF	80	C2-C10-NB	110	C6-C18-NB	140	C10-C17-NB	170	C17-C18-NB
21	C7-NB-RF	51	C17-NB-RF	81	C2-C10-RF	111	C6-C18-RF	141	C10-C17-RF	171	C17-C18-RF
22	C8-SVM-NB	52	C18-SVM-NB	82	C3-C6-SVM	112	C7-C8-SVM	142	C11-C16-SVM	172	C17-C19-SVM
23	C8-SVM-RF	53	C18-SVM-RF	83	C3-C6-NB	113	C7-C8-NB	143	C11-C16-NB	173	C17-C19-NB
24	C8-NB-RF	54	C18-NB-RF	84	C3-C6-RF	114	C7-C8-RF	144	C11-C16-RF	174	C17-C19-RF
25	C9-SVM-NB	55	C19-SVM-NB	85	C3-C9-SVM	115	C7-C12-SVM	145	C11-C18-SVM	175	C18-C20-SVM
26	C9-SVM-RF	56	C19-SVM-RF	86	C3-C9-NB	116	C7-C12-NB	146	C11-C18-NB	176	C18-C20-NB
27	C9-NB-RF	57	C19-NB-RF	87	C3-C9-RF	117	C7-C12-RF	147	C11-C18-RF	177	C18-C20-RF
28	C10-SVM-NB	58	C20-SVM-NB	88	C3-C11-SVM	118	C8-C9-SVM	148	C12-C13-SVM	178	C19-C20-SVM
29	C10-SVM-RF	59	C20-SVM-RF	89	C3-C11-NB	119	C8-C9-NB	149	C12-C13-NB	179	C19-C20-NB
30	C10-NB-RF	60	C20-NB-RF	90	C3-C11-RF	120	C8-C9-RF	150	C12-C13-RF	180	C19-C20-RF

is significant as it introduces a new topic within EC with the potential to help other EC research.

#### A. Future Work

The experiment we have conducted is only a small part of cross-language classification research since it only tested on the Danish and English language, a few preprocessing methods, and three classification algorithms. Therefore it is necessary to make similar experiments, e.g. on languages other than Danish and English in order to validate our hypothesis. Researching the cross-language effectiveness of other preprocessors and classifiers is also a possible continuation of our work. It will also be worth testing the differences between languages with different alphabets and/or structure, especially Latin-based and non-Latin-based languages. The framework described in Section IV can serve as a guide for comparing EC methods between languages. Whether languages have impact on the effectiveness of preprocessing and classification methods is still an open problem, that can be tested using other languages, preprocessing methods, classification algorithms, and/or data sets. One possible data set to use would be the SemEval-2019 data set<sup>6</sup>, which is used for a semantic evaluation workshop.

#### REFERENCES

- [1] M. V. Mäntylä, D. Graziotin, and M. Kuutila, "The evolution of sentiment analysis—a review of research topics, venues, and top cited papers," *Computer Science Review*, vol. 27, pp. 16 – 32, 2018. doi: 10.1016/j.cosrev.2017.10.002
- [2] R. Plutchik, "The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice," *American Scientist*, vol. 89, no. 4, pp. 344–350, 2001. doi: 10.1511/2001.4.344
- [3] G. Angiani *et al.*, "A comparison between preprocessing techniques for sentiment analysis in twitter," in *KDWeb*, 2016. doi: 10.1007/978-3-319-67008-9\_31
- [4] W. Wang, L. Chen, K. Thirunarayan, and A. P. Sheth, "Harnessing twitter "big data" for automatic emotion identification," in *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing*, Sep. 2012. doi: 10.1109/SocialCom-PASSAT.2012.119 pp. 587–592.
- [5] A. Balahur, "Sentiment analysis in social media texts," in *WASSA@NAACL-HLT*, 2013. doi: 10.1.1.310.4764
- [6] B. Gokulakrishnan, P. Priyanthan, T. Ragavan, N. Prasath, and A. Perera, "Opinion mining and sentiment analysis on a twitter data stream," in *International Conference on Advances in ICT for Emerging Regions (ICTer2012)*, Dec 2012. doi: 10.1109/ICTer.2012.6423033 pp. 182–188.
- [7] V. K. Jain, S. Kumar, and S. L. Fernandes, "Extraction of emotions from multilingual text using intelligent text processing and computational linguistics," *Journal of Computational Science*, vol. 21, pp. 316 – 326, 2017. doi: 10.1016/j.jocs.2017.01.010
- [8] M. Asad, N. Afroz, L. Dey, R. P. D. Nath, and M. A. Azim, "Introducing active learning on text to emotion analyzer," in *2014 17th International Conference on Computer and Information Technology (ICCIT)*, Dec 2014. doi: 10.1109/ICCITechn.2014.7073079 pp. 35–40.
- [9] J. R. Quinlan, "Induction of decision trees," *MACH. LEARN*, vol. 1, pp. 81–106, 1986. doi: 10.1007/BF00116251
- [10] C. Cortes and V. Vapnik, "Support-vector networks," in *Machine Learning*, 1995. doi: 10.1007/BF00994018 pp. 273–297.
- [11] J. C. Platt, "Sequential minimal optimization: A fast algorithm for training support vector machines," in *Advances in Kernel Methods-Support Vector Learning*, 1999.
- [12] T. Ho, "Random decision forests," in *Document Analysis and Recognition, International Conference on*, vol. 1, 09 1995. doi: 10.1109/ICDAR.1995.598994. ISBN 0-8186-7128-9 pp. 278 – 282 vol.1.
- [13] G. F. Cooper and E. HERSKOVITS, "A bayesian method for the induction of probabilistic networks from data," in *MACHINE LEARNING*, 1992. doi: 10.1007/BF00994110 pp. 309–347.
- [14] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bulletin*, vol. 1, no. 6, pp. 80–83, 1945.

<sup>6</sup><http://alt.qcri.org/semeval2019/>

# Signature analysis system using a convolutional neural network

Alicja Winnicka, Karolina Kęsik

Dawid Połap

Institute of Mathematics

Silesian University of Technology

Kaszubska 23, 44-100 Gliwice, Poland

Email: Alicja.Lidia.Winnicka@gmail.com, Karola.Ksk@gmail.com

Dawid.Polap@polsl.pl

**Abstract**—Identity verification using biometric methods has been used for many years. A special case is a handwritten signature made on a digital device or piece of paper. For the digital analysis and verification of its authenticity, special methods are needed. Unfortunately, this is a rather complicated task that quite often requires complex processing techniques. In this paper, we propose a system of signatures verification consisting of two stages. In the first one, a signature pattern is created. Thanks to this, the first attempt to verify identity takes place. In the case of approval, the second stage is followed by the processing of a graphic sample containing a signature by the convolutional neural network. The proposed technique has been described, tested and discussed due to its practical use.

## I. INTRODUCTION

USING signature, we can confirm our identity. This is particularly important in the case of signing contracts or receiving parcels from couriers. In each of these situations, we confirm something with signature. Of course, such a signature may differ each time we use it. The reason is his elaboration, which means that the more we sign, the more stable it will be. Lawyers, politicians or office workers who often sign on different documents will have a stable and permanent signature very quickly. Consequently, such a signature is very important as an element of our identity.

Unfortunately, there are situations where such a signature is forged for fraud purposes. To prevent this, it is worth having a signature, which minimizes the possibility of counterfeiting. This is possible thanks to a much smoother writing process, pen pressure, or runtime. All these features are often used in the authentication process. In practice, such verification is not an easy thing, which is why the complicated methods of artificial intelligence are often used.

An important element of verification process is the classifier, which makes decisions with a certain probability. The most popular are artificial neural networks which are inspired by the mechanisms occurring in the brain. An important element of scientific research is the improvement and design of new solutions that may later be used in biometrics [1]–[6]. One of the last achievements in this field are papers on the interpretation of the signature in numerical form and the use of so-processed samples in the training process [7], [8]. Interesting approach

was presented in [9], where the authors described a technique that uses a single record with a signature.

The cited works show great effectiveness, however, it is worth paying attention to the mechanism of such software. Quite a frequent mechanism is processing not on the device, but in the computing cloud [10]. Another thing is the front-end of the software and the ways in which the application is displayed to the user [11], [12]. Conducted research is not only related to the signature but also to other elements that can confirm our identity, an example of which are the fingerprint and iris of the eye [13]–[15].

In this paper, we present a system for signature verification based on two stages based on image processing and convolutional neural networks.

## II. SIGNATURE PROCESSING

Each graphic sample containing a signature should be processed. The main reason is the inclination of the signature relative to the straight line. In order to analyze or to compare two signatures, both samples should be arranged on the same straight line without any inclination.

The slope of the signature can be calculated using the linear approximation for the given set in the discrete form. The signature is a graphic file that should be saved in numerical form. The first step is to binarize the image. Each pixel in the image is described in the RGB color system (*Red-Green-Blue*), so the binarization will mean replacing all colors with white or black one using the following equation

$$\begin{cases} \frac{\sum_{i \in \{R,G,B\}} i(p)}{3} < \left\lfloor \frac{255}{2} \right\rfloor & \text{then } R(p) = G(p) = B(p) = 0 \\ \frac{\sum_{i \in \{R,G,B\}} i(p)}{3} > \left\lfloor \frac{255}{2} \right\rfloor & \text{then } R(p) = G(p) = B(p) = 255 \end{cases} \quad (1)$$

where functions  $R(\cdot)$ ,  $G(\cdot)$  and  $B(\cdot)$  are a color component of pixel  $p$ . The value on the right of the inequality means the total value of the center of the color range in the RGB model.

As already mentioned, the image after the binarization process consists of two colors – black and white, where

the black pixels represent the signature. Each pixel can be interpreted as a coordinate  $(x, y)$ . Taking these points, we have a set of points on the Cartesian plane given discreetly. On this basis, it is possible to calculate the slope of the signature using the linear function equation in the following form

$$f(x) = y = a_0 + a_1x \Rightarrow a_0 = \tan(\alpha) \Rightarrow \alpha = \arctan(a_0). \quad (2)$$

It is easy to see that having a coefficient  $a_0$ , the slope can be calculated. However, the value of  $a_0$  must be calculated. Suppose that the set of points has  $n$  elements, then we are looking for such a function  $f(x)$ , for which the following condition will occur

$$f(x_i) = y_i, \quad i = 0, \dots, n-1. \quad (3)$$

Assume that for a set of points  $\{(x_i, y_i)\}$  where  $i \in \{0, n-1\}$ , a function  $S(a_0, a_1)$  will be presented as

$$S(a_0, a_1) = \sum_{i=0}^{n-1} (y_i - a_0 - a_1x_i)^2 \quad (4)$$

Therefore, the system of normal equations has the following form

$$\frac{\partial S(a_0, a_1)}{\partial a_0} = \sum_{i=0}^{n-1} (y_i - a_0 - a_1x_i)(-1) = 0 \quad (5)$$

$$\frac{\partial S(a_0, a_1)}{\partial a_1} = \sum_{i=0}^{n-1} (y_i - a_0 - a_1x_i)(-x_i) = 0 \quad (6)$$

By grouping the above two equations, we get

$$a_0n + a_1 \sum_{i=0}^{n-1} x_i = \sum_{i=0}^{n-1} y_i \quad (7)$$

$$a_0 \sum_{i=0}^{n-1} x_i + a_1 \sum_{i=0}^{n-1} x_i + a_1 \sum_{i=0}^{n-1} x_i^2 = \sum_{i=0}^{n-1} x_i y_i. \quad (8)$$

The above system of equations is linear, so it can be saved in a simpler form as

$$X \cdot A = Y, \quad (9)$$

where  $A, X, Y$  are a matrices defined as

$$A = \begin{pmatrix} a_0 \\ a_1 \end{pmatrix}, X = \begin{pmatrix} \sum_{i=0}^{n-1} 1 & \sum_{i=0}^{n-1} x_i \\ \sum_{i=0}^{n-1} x_i & \sum_{i=0}^{n-1} x_i^2 \end{pmatrix}, Y = \begin{pmatrix} \sum_{i=0}^{n-1} y_i \\ \sum_{i=0}^{n-1} x_i y_i \end{pmatrix} \quad (10)$$

Finally, searched coefficients can be obtained by

$$A = X^{-1} \cdot Y. \quad (11)$$

In this way, we obtain the coefficients of the approximated linear function, and thus using the equation (2), it is possible to find the slope of the signature for which the image with the signature should be rotated.

Having a processed and rotated image, we put together several signatures belonging to the same person. The imposition of images consists in creating matrix with a dimension adequate to the samples (if the samples are of different sizes, they should be normalized to the same one). This matrix should be filled with 0 (which is understood as a white pixel). Then, for each image, pixels are checked. If there is a black pixel at a given position in the image, then the value in this matrix is increased by 1.

Such a matrix allows us to create patterns. The higher the value of a given matrix element, the more often the pixel appears and can be treated as a feature.

The smallest values should be replaced with zeros. The selection of this value depends on the number of samples used in the process of its creation. As part of the experiments to be carried out, the optimal value was determined as  $\frac{n}{2}$  or  $\frac{n-1}{2}$ .

Such a matrix can be applied to a new, processed signature. In this case, we calculate the number of black pixels of the signature that are on the positions in the matrix (where the elements are different from zero). This allows us to calculate the percentage coverage of features.

### III. CONVOLUTIONAL NEURAL NETWORK

The previous stage allowed the creation of a technique that gives the percentage quality of coverage of the main features. Unfortunately, it does not allow the verification process itself. For this purpose, we use Convolutional Neural Network (CNN) [16], which are a mathematical model of action having place in the primary cortex. These structure take the image at the entrance, and at the output they return the class to which the input belongs with a certain probability. Structure construction can be described using three layers. The first two layers are used to feature extraction and are called convolutional and pooling. The first type processes the image using a certain filter  $\omega$  defined as a matrix and a step  $S$  through which this matrix will be moved. The next layer is called pooling which reduces the size of the image using the selected function. If the image is to be reduced to  $t$  times, a matrix of size  $t \times t$  is created, in which only one pixel is selected and this matrix is shifted in the image, resulting in a reduced image. Next, a classic neural network is created that forms the last layer of the network.

The structure itself is quite simple in its model, however, it must be added that the layers are connected to each other thanks to synapses burdened with weights. At the initial stage of creating networks, they are generated in a random way. Using the training algorithm, they are modified due to input data. The most commonly used algorithm is Adaptive Moment Estimation (*Adam*). The algorithm consists in calculating the mean values  $m$  of the gradient and the second momentum (variance)  $v$  in each iteration  $t$ . Let us assume that  $w^{(t)}$  will be understood as a parameters and  $L^{(t)}(\cdot)$  as a loss function. Formally, the equations for these values are as follows

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t, \quad (12)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2, \quad (13)$$

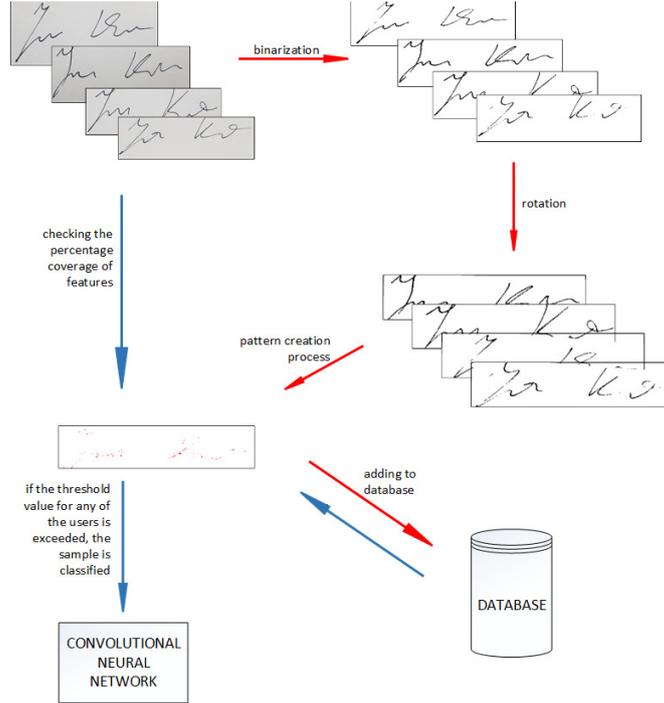


Figure 1: Red arrows indicate the preparation process and blue arrows verification process.

Layer	Output Shape
Convolutional	(None,148,148,32)
Activation	(None,148,148,32)
MaxPooling	(None,74,74,32)
Convolutional	(None,72,72,32)
Activation	(None,72,72,32)
MaxPooling	(None,36,36,32)
Convolutional	(None,34,34,64)
Activation	(None,34,34,64)
MaxPooling	(None,17,17,64)
Flatten	(None,18496)
Dense	(None,64)
Activation	(None,64)
Dropout	(None,64)
Dense	(None,2)
Activation	(None,2)

Table I: Convolutional neural network architecture.

where  $\beta_1$  and  $\beta_2$  are decay coefficients, which values are close to 1. The correction for a given moment is defined as

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}, \quad (14)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t}, \quad (15)$$

which are used to update the value

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} \hat{m}_t, \quad (16)$$

where  $\epsilon$  is a constant, small value used to prevent dividing by 0 and  $\eta$  is the learning rate.

#### IV. SIGNATURE VERIFICATION MODEL

The proposed system consists of image processing or normalization of the sample, and then using it in two stages. The first one is to create a matrix or use it to check the percentage of coverage of features. If the obtained value exceeds the threshold value, then this sample is classified by the convolutional neural network. The graphical illustration of the model is shown in Fig. 1. This action reduces the number of operations performed by possibly rejecting the sample due to the feature matrix.

Assuming that the system should enable the identity verification of several people, the signature is analyzed in relation to all matrices in the database. If for any of them, the percentage threshold is exceeded, it is classified by the network. In case the matrix and network return different results, the system will not be able to clearly identify the owner of the sample.

#### V. EXPERIMENTS

For testing purposes, a small signature database of two people was created, consisting of 50 samples (25 per person). In addition, 20 samples (10 for each person) of fake signatures were created (tried to imitate other's signatures). In the classifier learning process, all samples were normalized to the dimension  $150 \times 150$  pixels. Used architecture of CNN is presented in Tab. I.

The features matrix was tested for different values, and the best efficiency (in an empirical way) was obtained for a value equal to  $\frac{n-1}{2}$ . Classifier was trained 10 epochs using 70% : 30% of samples (training to validation number of samples). The history of training is shown in Fig. 2

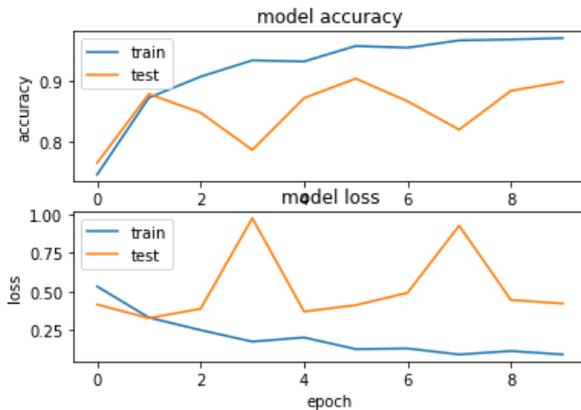


Figure 2: Graphs of training history using 50 samples (35:15 training : validation).

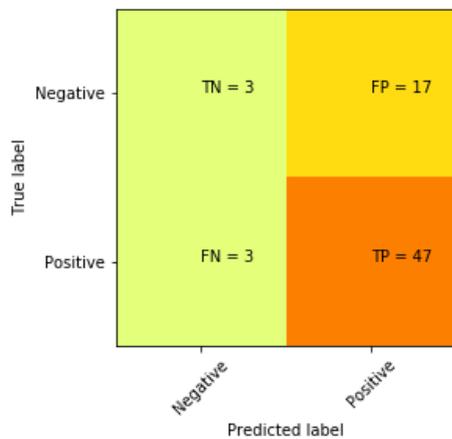


Figure 3: Confusion matrix for average results contained from 10 attempts.

The effectiveness of the classification for two people was achieved at 94%, which is a very good result considering the number of samples in the training process. It is worth noting that after training the classifier, fake samples were used in order to verify the operation of the proposed system. 11 signatures were rejected at the level of the features matrix and the remaining ones were classified by the classifier. The average probability of belonging to these people was in the range 33–90%, what is a good result. The optimistic approach is the result of the fact that the original samples were classified above 83%, what allows to reject counterfeit signatures except for selected images. The network was trained ten times in order to obtain an average classification value, which was achieved at 91%, the average results of classification was presented in confusion matrix in Fig. 3.

## VI. CONCLUSION

The described method of identity verification based on the signature indicates high efficiency. The experiments were

carried out on the basis of a total sum of samples equal to 50, which is quite a small amount. It is worth noting that adding verification using the matrix of features allowed to reduce the operations performed using CNN because it rejected over half of suspicious samples before the verification stage. This solution indicates the possibility of obtaining better efficiency using more extensive techniques for creating matrix pattern using other features.

## ACKNOWLEDGMENTS

Authors acknowledge contribution to this project to the Diamond Grant No. 0080/DIA/2016/45 funded by the Polish Ministry of Science and Higher Education.

## REFERENCES

- [1] I. Rocco, R. Arandjelovic, and J. Sivic, "Convolutional neural network architecture for geometric matching," *IEEE transactions on pattern analysis and machine intelligence*, 2018.
- [2] V. Nourani, S. Mousavi, D. Dabrowska, and F. Sadikoglu, "Conjunction of radial basis function interpolator and artificial intelligence models for time-space modeling of contaminant transport in porous media," *Journal of hydrology*, vol. 548, pp. 569–587, 2017.
- [3] D. Dąbrowska, R. Kucharski, and A. J. Witkowski, "The representativity index of a simple monitoring network with regular theoretical shapes and its practical application for the existing groundwater monitoring network of the tychy-urbanowice landfills, poland," *Environmental Earth Sciences*, vol. 75, no. 9, p. 749, 2016.
- [4] A. Venčkauskas, R. Damaševičius, R. Marcinkevičius, and A. Karpavičius, "Problems of authorship identification of the national language electronic discourse," in *International Conference on Information and Software Technologies*. Springer, 2015, pp. 415–432.
- [5] R. Damaševičius, R. Maskeliūnas, E. Kazanavičius, and M. Woźniak, "Combining cryptography with eeg biometrics," *Computational intelligence and neuroscience*, vol. 2018, 2018.
- [6] R. Damaševičius, R. Maskeliūnas, A. Venčkauskas, and M. Woźniak, "Smartphone user identity verification using gait characteristics," *Symmetry*, vol. 8, no. 10, p. 100, 2016.
- [7] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, and J. Ortega-Garcia, "Exploring recurrent neural networks for on-line handwritten signature biometrics," *IEEE Access*, vol. 6, no. 5128–5138, pp. 1–7, 2018.
- [8] M. Elhoseny, A. Nabil, A. E. Hassanien, and D. Oliva, "Hybrid rough neural network model for signature recognition," in *Advances in Soft Computing and Machine Learning in Image Processing*. Springer, 2018, pp. 295–318.
- [9] M. Diaz, A. Fischer, M. A. Ferrer, and R. Plamondon, "Dynamic signature verification system based on one real signature," *IEEE Transactions on Cybernetics*, vol. 48, no. 1, pp. 228–239, 2018.
- [10] G. L. Masala, P. Ruiu, and E. Grosso, "Biometric authentication and data security in cloud computing," in *Computer and Network Security Essentials*. Springer, 2018, pp. 337–353.
- [11] Z. Sroczynski, "Actiontracking for multi-platform mobile applications," in *Computer Science On-line Conference*. Springer, 2017, pp. 339–348.
- [12] A. Bier and Z. Sroczynski, "Towards semantic search for mathematical notation," in *2018 Federated Conference on Computer Science and Information Systems (FedCSIS)*. IEEE, 2018, pp. 465–469.
- [13] N. Merhav, "Ensemble performance of biometric authentication systems based on secret key generation," *IEEE Transactions on Information Theory*, 2018.
- [14] K. Zhou and J. Ren, "Passbio: Privacy-preserving user-centric biometric authentication," *IEEE Transactions on Information Forensics and Security*, 2018.
- [15] P. Gupta and P. Gupta, "Multibiometric authentication system using slap fingerprints, palm dorsal vein, and hand geometry," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 12, pp. 9777–9784, 2018.
- [16] H. Huang, C. Wang, and B. Dong, "Nostalgic adam: Weighing more of the past gradients when designing the adaptive learning rate," *arXiv preprint arXiv:1805.07557*, 2018.

# 12<sup>th</sup> International Symposium on Multimedia Applications and Processing

**O**RGANIZED by Software Engineering Department, Faculty of Automation, Computers and Electronics, University of Craiova, Romania “Multimedia Applications Development” Research Centre

## BACKGROUND AND GOALS

Multimedia information has become ubiquitous on the web, creating new challenges for indexing, access, search and retrieval. Recent advances in pervasive computers, networks, telecommunications, and information technology, along with the proliferation of multimedia mobile devices—such as laptops, iPods, personal digital assistants (PDA), and cellular telephones—have stimulated the development of intelligent pervasive multimedia applications. These key technologies are creating a multimedia revolution that will have significant impact across a wide spectrum of consumer, business, healthcare, educational and governmental domains. Yet many challenges remain, especially when it comes to efficiently indexing, mining, querying, searching, retrieving, displaying and interacting with multimedia data.

The Multimedia—Processing and Applications 2019 (MMAAP 2019) Symposium addresses several themes related to theory and practice within multimedia domain. The enormous interest in multimedia from many activity areas (medicine, entertainment, education) led researchers and industry to make a continuous effort to create new, innovative multimedia algorithms and applications.

As a result the conference goal is to bring together researchers, engineers, developers and practitioners in order to communicate their newest and original contributions. The key objective of the MMAAP conference is to gather results from academia and industry partners working in all subfields of multimedia: content design, development, authoring and evaluation, systems/tools oriented research and development. We are also interested in looking at service architectures, protocols, and standards for multimedia communications—including middleware—along with the related security issues, such as secure multimedia information sharing. Finally, we encourage submissions describing work on novel applications that exploit the unique set of advantages offered by multimedia computing techniques, including home-networked entertainment and games. However, innovative contributions that don't exactly fit into these areas will also be considered because they might be of benefit to conference attendees.

## CALL FOR PAPERS

MMAAP 2019 is a major forum for researchers and practitioners from academia, industry, and government to present, discuss, and exchange ideas that address real-world problems with real-world solutions.

The MMAAP 2019 Symposium welcomes submissions of original papers concerning all aspects of multimedia domain ranging from concepts and theoretical developments to advanced technologies and innovative applications. MMAAP 2019 invites original previously unpublished contributions that are not submitted concurrently to a journal or another conference. Papers acceptance and publication will be judged based on their relevance to the symposium theme, clarity of presentation, originality and accuracy of results and proposed solutions.

## TOPICS

- Audio, Image and Video Processing
- Animation, Virtual Reality, 3D and Stereo Imaging
- Big Data Science and Multimedia Systems
- Cloud Computing and Multimedia Applications
- Machine Learning, Information Retrieval in Multimedia Applications
- Data Mining, Warehousing and Knowledge Extraction
- Multimedia File Systems and Databases: Indexing, Recognition and Retrieval
- Multimedia in Internet and Web Based Systems
- E-Learning, E-Commerce and E-Society Applications
- Human Computer Interaction and Interfaces in Multimedia Applications
- Multimedia in Medical Applications and Computational biology
- Entertainment, Personalized Systems and Games
- Security in Multimedia Applications: Authentication and Watermarking
- Distributed Multimedia Systems
- Network and Operating System Support for Multimedia
- Mobile Network Architecture and Fuzzy Logic Systems
- Intelligent Multimedia Network Applications
- Future Trends in Computing System Technologies and Applications
- Trends in Processing Multimedia Information
- Multimedia Ontology and Perception for Multimedia Users

#### BEST PAPER AWARD

A best paper award will be made for work of high quality presented at the MMAP Symposium. The technical committee in conjunction with the organizing/steering committee will decide on the qualifying papers. Award comprises a certificate for the authors and will be announced on time of conference.

#### STEERING COMMITTEE

- **Amy Neustein**, Boston University, USA, Editor of Speech Technology
- **Lakhmi C. Jain**, University of South Australia and University of Canberra, Australia
- **Zurada, Jacek**, University of Louisville, United States
- **Ioannis Pitas**, University of Thessaloniki, Greece
- **Costin Badica**, University of Craiova, Romania
- **Borko Furht**, Florida Atlantic University, USA
- **Harald Kosch**, University of Passau, Germany
- **Vladimir Uskov**, Bradley University, USA
- **Thomas M. Deserno**, Aachen University, Germany

#### HONORARY CHAIR

- **Dumitru Dan Burdescu**, University of Craiova, Romania

#### GENERAL CO-CHAIRS

- **Adriana Schiopoiu Burlea**, University of Craiova, Romania
- **Marius Brezovan**, University of Craiova, Romania
- **Marcin Woźniak**, Institute of Mathematics, Silesian University of Technology, Poland

#### PUBLICITY CHAIR

- **Amelia Badica**, University of Craiova, Romania
- **Milan Simic**, RMIT University, School of Engineering, Australia

#### ORGANIZING

- **Dumitru Dan Burdescu**, University of Craiova, Romania
- **Costin Badica**, University of Craiova, Romania
- **Marius Brezovan**, University of Craiova, Romania
- **Adriana Schiopoiu Burlea**, University of Craiova, Romania
- **Liana Stanescu**, University of Craiova, Romania
- **Cristian Marian Mihaescu**, University of Craiova, Romania

#### PROGRAM COMMITTEE

- **Azevedo, Ana**, CEOS.PP-ISCAP/IPP, Portugal
- **Badica, Amelia**, University of Craiova, Romania
- **Burlea Schiopoiu, Adriana**, University of Craiova, Romania
- **Cano, Alberto**, Virginia Commonwealth University, United States
- **Cordeiro, Jose**, EST Setúbal/I.P.S.

- **Cretu, Vladimir**, Politehnica University of Timisoara, Romania
- **Debono, Carl James**, University of Malta, Malta
- **Fabijańska, Anna**, Lodz University of Technology, Poland - Institute of Applied Computer Science, Poland
- **Fomichov, Vladimir**, National Research University Higher School of Economics, Moscow, Russia., Russia
- **Giurca, Adrian**, Brandenburg University of Technology, Germany
- **Grosu, Daniel**, Wayne State University, United States
- **Kabranov, Ognian**, Cisco Systems, United States
- **Keswani, Dr. Bright**, Suresh Gyan Vihar University, Mahal, Jagatpura, Jaipur
- **Korzhih, Valery**, State University of Telecommunications, Russia
- **Kotenko, Igor**, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Science, Russia
- **Logofatu, Bogdan**, University of Bucharest, Romania
- **Mangioni, Giuseppe**, DIEEI - University of Catania, Italy
- **Marghitu, Daniela**, Auburn University
- **Mihaescu, Cristian**, University of Craiova, Reunion
- **Mocanu, Mihai**, University of Craiova, Romania
- **Ohzeki, Kazuo**, Professor Emeritus at Shibaura Institute of Technology, Japan
- **Pohl, Daniel**, Intel, Germany
- **Popescu, Dan**, CSIRO, Sydney, Australia, Australia
- **Popescu, Daniela E.**, Integrated IT Management Service, University of Oradea
- **Querini, Marco**, Department of Civil Engineering and Computer Science Engineering
- **Radulescu, Florin**, University "Politehnica" of Bucharest
- **Romansky, Radi**, Professor at Technical University of Sofia
- **RUTKAUSKIENE, Danguole**, Kaunas University of Technology
- **Salem, Abdel-Badeeh M.**, Ain Shams University, Egypt
- **Sari, Riri Fitri**, University of Indonesia, Indonesia
- **Scherer, Rafał**, Częstochowa University of Technology, Poland
- **Sousa Pinto, Agostinho**, Instituto Politécnico do Porto, Portugal
- **Stanescu, Liana**, University of Craiova, Romania
- **Stoicu-Tivadar, Vasile**, University Politehnica Timisoara
- **Trausan-Matu, Stefan**, Politehnica University of Bucharest, Romania
- **Trzcielinski, Stefan**, Poznan University of Technology, Poland
- **Tsihrintzis, George**, University of Piraeus, Greece
- **Tudoroiu, Nicolae**, John Abbott College, Canada
- **Vega-Rodríguez, Miguel A.**, University of Extremadura, Spain
- **Virvou, Maria**, University of Piraeus, Greece
- **Watanabe, Toyohide**, University of Nagoya

# Creating See-Around Scenes using Panorama Stitching

Saja Alferidah  
King Faisal University  
Saudi Arabia, Alahsa  
Email:saja.alferidah@gmail.com

Nora A. Alkhaldi  
King Faisal University  
Saudi Arabia, Alahsa  
Email: nalkhaldi@kfu.edu.sa

**Abstract**—Image stitching refers to the process of combining multiple images of the same scene to produce a single high-resolution image, known as panorama stitching. The aim of this paper is to produce a high-quality stitched panorama image with less computation time. This is achieved by proposing four combinations of algorithms. First combination includes FAST corner detector, Brute Force K-Nearest Neighbor (KNN) and Random Sample Consensus (RANSAC). Second combination includes FAST, Brute Force (KNN) and Progressive Sample Consensus (PROSAC). Third combination includes ORB, Brute Force (KNN) and RANSAC. Fourth combination contains ORB, Brute Force (KNN) and PROSAC. Next, each combination involves a calculation of Transformation Matrix. The results demonstrated that the fourth combination produced a panoramic image with the highest performance and better quality compared to other combinations. The processing time is reduced by 67% for the third combination and by 68% for the fourth combination compared to stat-of-the-art.

## I. INTRODUCTION

THE STUDY of panoramic imaging is one of the advanced research topics in the field of computer vision, graphics and image processing [1]. Panorama Stitching is defined when two or more images of the same scene are taken by rotating a camera about its axis. As a result of this process a wider panorama image is created by overlapping the common contents of each component image [2]. In 1997, Szelinski and Shum defined creating a larger panorama image as the integration and overlapping the common contents of two or more images of the same scene by rotating the camera about its axis. In 2017, Wand et al. defined panorama stitching as taking multiple images with an overlapping area and stitching them together into a single wide image [3][4]. In 2015, Hee-kyeong Jeon et al. classified the panorama stitching process as the three core steps of detecting features, matching them, and stitching [2]. Early panorama images were created by sliding a slit-shaped aperture across a photographic film. The digital approach of today extracts thin, vertical strips of pixels from the frames of a sequence captured by a translating video camera. The resulting image is considered as multi-viewpoint (or multi-perspective), because different strips of the image are captured from multiple viewpoints [4]. Strip panoramas are created from a translating camera with many variants, such as "pushbroom panoramas" [5], "adaptive manifolds" [6], and "x-slit" images [7]. Contrary to the hardware-based approach,

many researchers have explored the multi-perspective renderings of 3D models [8][9]. Yu and McMillan presented a model that describes a multi-perspective camera [10]. Panoramic image stitching is used in a variety of environment, including gaming, virtual reality, virtual museums, and map applications [11]. Microsoft Research, for example, is spending on research projects featuring panorama stitching techniques, and many algorithms are designed to efficiently facilitating the creation of panoramic images through stitching [12][13].

## II. BACKGROUND

Most researchers classify panorama stitching as either a direct technique or a feature-based technique [11][14]. The direct technique compares pixel to pixel between images and the feature-based technique compares all features within each image [14]. This paper applies the feature-based technique as it is more advanced, faster, and flexible when compared to the direct technique. Producing a panorama stitching for two or more images of the same object is divided into three steps. First, the process discovers the points of interest between several images (the keypoints) and extract vector features around each of these points of interest (the descriptors). Second, identifies the matching lines between several images using the extracted features after that match the correct features and remove incorrect features. Third, find the transformation matrix that satisfies matching with the other keypoints, and use this transformation to align the two images before merging. Panorama stitching is considered through two perspectives. The first is camera rotation, where images are acquired with the camera positioned at the same point while being rotated to provide multiple views of the same object. The second perspective is camera translation, where the camera is not fixed at the same position but is moved through a linear translation to capture the second image. This paper focuses on the second prospective where two images are taken for the same scene and with a slight linear displacement.

Consider a car moving towards an intersection with a large building on the corner obstructing its view. If an image is taken from a point ahead of its current position and stitched with another image at its current position, such that the integrated image shows the two overlapped as a semi-transparent view, then this image can enable drivers to have a partial view of the scene behind the building. This work helps to create a vision

effect around the image, Figure 1, graphically explains this scenario. The first camera (Camera 1) captures one image of

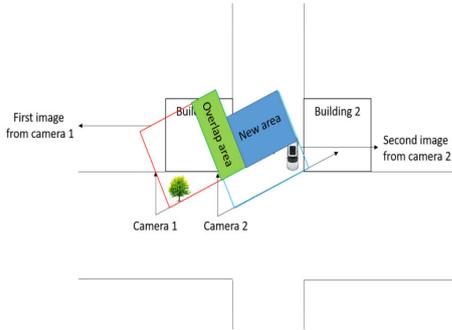


Fig. 1: The panorama stitching problem illustrated

building 1, which is shown as a red square. The blue square is the image captured from the second camera (Camera 2) that captures part of building 1 and part of building 2. The green rectangle represents the overlapping area between the two images and the blue area behind the green rectangle is the portion obscured by the building. This paper employs seven techniques, which are combined into four hybrid models, as shown in Figure 2, to create panoramic images. The four

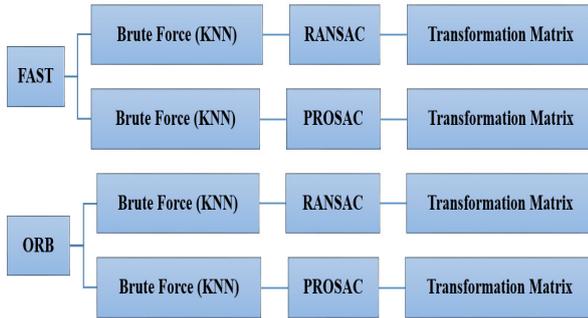


Fig. 2: The selected solution for four hybrid approaches

combinations include: (1) *FAST*, Brute Force (KNN) and *RANSAC*; (2) *FAST*, Brute Force (KNN) and *PROSAC*; (3) *ORB*, Brute Force (KNN) and *RANSAC*; (4) *ORB*, Brute Force (KNN) and *PROSAC*. Then each combination is followed by the calculation of a Transformation Matrix. The results of these models are compared to the model proposed in [2]. Basically, the model in [2] utilized *ORB*, Hamming distance, *PROSAC* and the Transformation Matrix to produce a stitched image. This model will be referred to as the fifth combination here. Next sections will discuss the seven implemented techniques.

#### A. *FAST*

The *FAST* technique is a high-speed corner detector method [15] defined by having a pixel A surrounded by a sufficient quantity of neighborhood pixels with a different grayscale value. In this scenario, the pixel A is recognized as

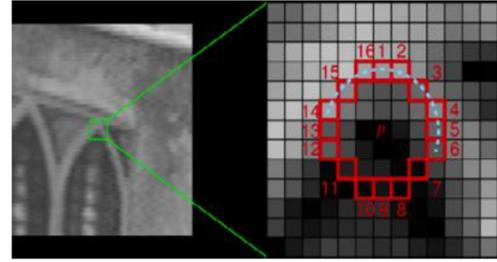


Fig. 3: *FAST* Feature Point Detection [15]

a *FAST* corner and applies to grayscale images. The *FAST* neighborhood must contain enough pixels with values greater than, less than or similar to that of pixel A. We choose an arbitrary pixel as the center to establish a circular area, to be considered as the pixel point's neighborhood [16]. As shown in Figure 3, a discrete circle of radius 3 with pixel p as the central pixel has neighborhood pixels labeled 1 to 16. If pixel 16 has sequential n pixels that satisfy the equation [16]

$$|I_x - I_p| > t, \quad (1)$$

then, we consider p as a candidate feature point, t is a given threshold value,  $I_x$  is the gray value of the sequential n pixel, and  $I_p$  is the gray value of pixel p, [17]. For features extraction and descriptors computation in the first and second hybrid combinations based on using *FAST*, the Binary Robust Invariant Scalable Keypoints (*BRISK*) algorithm is incorporated, [23], because *FAST* can only detect corner features but dose not compute the descriptors. Therefore, this paper uses *BRISK* descriptor with *FAST* keypoints.

#### B. *BRISK*

*BRISK* is a binary descriptor that calculate the weighted Gaussian average over selected points near the keypoint [23]. For specific pairs of Gaussian windows *BRISK* compare values that could be either a 1 or a 0 depending on which window in pair was greater [23]. *BRISK* descriptor applies the sampling pattern around the keypoints [23]. The sampling pattern rotated  $\alpha$  angle around the keypoint k. The  $\alpha$  is calculated by [23]:

$$\alpha = \arctan 2(g_y, g_x), \quad (2)$$

where  $g_x$  and  $g_y$  are the gradients sum. The bit vector descriptor  $d_k$  is collected by execute for all point pairs the short distance intensity comparisons [23].

$$(p_i^\alpha, \sigma_i) \in S, \quad (3)$$

such as every bit b corresponds to:

$$b = \begin{cases} 1, & I(p_j^\alpha, \sigma_j) > I(p_i^\alpha, \sigma_i) \\ 0, & \text{otherwise} \end{cases}, \quad (4)$$

$$\forall (p_i^\alpha, p_j^\alpha) \in S, \quad (5)$$

where  $I(p_i^\alpha, \sigma_i)$  is gray intensity after rotated  $\alpha$  angle around the keypoint k and S is gray intensity for the short distance

pairs set. At the end, *BRISK* uses a deterministic sampling pattern introduce a uniform sampling-point density [23].

### C. ORB

The *ORB* technique is based on improved *FAST* and the Binary Robust Independent Elementary Features *BRIEF* feature detector techniques to extract points of interest by using a binary string [18]. Since *FAST* and *BRIEF* process quickly, the *ORB* will also be fast [15]. While the *FAST* technique is not sensitive to noise and is highly reliable for identifying feature points, it does not provide an orientation. However, *ORB* incorporates orientation into *FAST* with the *oFAST* algorithm. The *BRIEF* approach finds descriptors around each feature point by using a binary coding method [19], which is simple and requires less memory compared to *SIFT* and *SURF* [19]. Consider  $p$  is a smoothed image patch defined on the size of  $S*S$  (Where  $S$  contains the coordinates of pixels) round feature points and a binary random selected test defined as  $\tau$ ,

$$\tau(p; x; y) = \begin{cases} 1, & p(x) < p(y) \\ 0, & \text{otherwise} \end{cases}, \quad (6)$$

where  $P(x)$  is the pixel intensity at the  $x = (u, v)^T$  point [3][20][16]. After filtering is performed, a set of points can uniquely identify one binary detection  $\tau$  [16]. Therefore, the features defined as a vector of  $n$  binary strings is the same as,

$$f_n(p) = \sum_{1 \leq i \leq n} 2^{i-1} \tau(p; x; y), \quad (7)$$

[3][20][16]. Since *BRIEF* is not scale invariant, *ORB* solves this issue by adding a direction into *BRIEF* by defining patch moments as [3][20][16]

$$m_{pq} = \sum_{x,y} x^p y^q I(x, y), \quad (8)$$

where  $p$  and  $q \in \{0, 1\}$  is binary selector for  $x$  and  $y$  direction and  $(x, y)$  is the position of the *FAST* feature point. The circular neighborhood radians are  $r, x, y \in [-r, r]$  [21][22], and the moment is reordered (centroid) as  $C$  [3][20][22], such that

$$C = \left( \frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right). \quad (9)$$

When assuming a center vector  $O$  to the centroid  $\vec{oc}$ , then the offset is defined as

$$\Theta = \arctan\left[\frac{m_{01}}{m_{10}}\right] = \arctan\left[\frac{\sum_{x,y} y I(x, y)}{\sum_{x,y} x I(x, y)}\right], \quad (10)$$

[3][20][16]. Therefore, *ORB* extracts the *BRIEF* descriptor based on the direction performed by Equation 6. The random *ORB* uses a greedy algorithm to find the random pixel block with low correlation and vector length equal to a 256-bit feature descriptor named *BRIEF*, [16], for which some previous research used a different type of test, such as the Gaussian distribution [3].

### D. K-Nearest Neighbor

One image matching algorithm is *KNN* that take a set of query points  $Q$  and set of references point  $R$  [24]. Then, check for each query point  $q \in Q$ , compute distance between  $q$  and all  $r \in R$ , sort the computed distance in list [24]. Finally, select  $K$  nearest reference points corresponding to  $k$  smallest distance [24]. A threshold ratio value is then checked to determine if it is a good matching point, which requires the process to loop until at least four matches are found to compute the Homography [25]. This paper uses the Brute Force matcher, a simple version of the *KNN*, to match the descriptors of the images.

### E. RANSAC

*RANSAC* is a robust technique used to estimate the Homography and remove outlier points randomly from images to provide good matches [11] and increase quality [2]. *RANSAC* randomly select a set of data required to calculate a mathematical model of data parameters [26][16]. Then, with an effective random sample [16], *RANSAC* uses a small number of points to estimate the model and check if it agrees with the remaining points by calculating their distance to the fitted model. *RANSAC* can be performed  $N$  times until a subset of the image is found with a good matching relationship [11].

### F. PROSAC

The *PROSAC* technique is used to remove outliers points progressively from images to obtain good matching results [26]. This algorithm performs the same steps as *RANSAC* gradually and not randomly, which reduces the required operation time and the number of repetition when the sufficient process of verification is completed [2]. Two problems need to be addressed in *PROSAC*, first is the growth function [26],

$$n = g(t), \quad (11)$$

which is defined as the set  $U_n$  of  $n$  (where  $U$  is set of features) arranged progressively and sampled after trials  $t$  are selected [19]. Second, *PROSAC* like *RANSAC* provides guarantees about the stopping criterion for the optimal solution, which must be found [26].

### G. Transformation Matrix

The transformation matrix defines  $x$  in an image  $B$  with  $x'$  as the final panorama image for calculating a new position of pixels [27][3],

$$x' \sim H_x, \quad (12)$$

where  $x$  is a new position of pixels in image  $B$ ,  $x'$  is a position of pixels in the final panorama,  $\sim$  finds the similarity up to scale, and  $H$  is a  $3 \times 3$  matrix that can be calculated using the Direct Linear Transform algorithm [27]

$$H = \begin{pmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{pmatrix}, \quad (13)$$

where points  $x$  and  $x'$  are defined as [27]

$$x_i = \begin{pmatrix} x_i \\ y_i \\ w_i \end{pmatrix}, x'_i = \begin{pmatrix} x'_i \\ y'_i \\ w'_i \end{pmatrix}, \quad (14)$$

where  $x_i, y_i$  is the keypoint position and  $w_i$  is set to 1. The final equation after subsequent transformation [27] becomes:

$$\begin{pmatrix} 0^T & -w'_i x_i^T & y'_i x_i^T \\ w'_i x_i^T & 0^T & -x'_i x_i^T \\ -y'_i x_i^T & x'_i x_i^T & 0^T \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix} = 0. \quad (15)$$

Two linearly independent equations. In addition, this can be written as [27]:

$$A \begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix} = 0. \quad (16)$$

We add two question on matrix A for each pair of points.

### III. EXPERIMENTAL RESULTS

The machine used in this work is Windows 10 with 64-bit operating system. The application uses two cameras with  $1373 \times 2382$  image resolution to capture sets of images for testing the panorama stitching algorithms. The experiments are performed using 15 different scenes, each with two captured images. Contrast Differences is used to evaluate the quality of the four hybrid combinations to determine if the stitched images are seamless as the seam is considered poor when it is visible.

#### A. Contrast Differences

Contrast differences are the variances in luminance between neighboring pixels that make them distinguishable [28]. In this paper, the differences in the contrast value between the stitched images and the original image check the quality of the four stitched images. Equation 17 shows the image contrast calculation  $I'_k$ , where  $I_k$  is the image in the vertical direction, is determined by:

$$I'_k(i, j) = I_k(i + 1, j) - I_k(i, j), \quad (17)$$

for  $1 \leq i < H$  and  $1 \leq j \leq L$ . To evaluate the quality between the original and the four stitched images, the area of the original image and the four stitched images is divided [28] as illustrated in Figure 4 with the stitched image  $I_{k,k+1}$  and the two halves of the overlapping area from the original images  $A$  and  $B$  and  $t_v$  is the horizontal translation.

The left half of the overlapping area is mainly contributed by the left half of the stitched image from  $A'$ . The right half of overlapping area is mainly contributed by the right half of the stitched image from  $B'$ . So, the contrast values of  $A$  and  $B$  are subtracted from contrast values of  $A'$  and  $B'$  to calculate the contrast difference values. The performance measures  $d_A$  and  $d_B$  for the  $A$  and  $B$  regions are then calculated as [28].

$$d_A = \sum_{i=1}^H \sum_{j=1}^{t_v} |I'_k(i, L_k - t_v + j) - I'_{k,k+1}(i, L_k - t_v + j)| \quad (18)$$

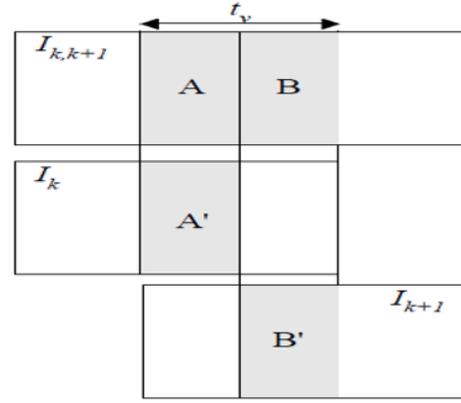


Fig. 4: Regions for comparison [28]

$$d_B = \sum_{i=1}^H \sum_{j=1}^{\frac{t_v}{2}} |I'_{k+1}(i, \frac{t_v}{2} + j) - I'_{k,k+1}(i, L_k - \frac{t_v}{2} + j)| \quad (19)$$

where the  $L$  and  $H$  are the width and height of the images, respectively [28]. Beside using Contrast Differences, this paper use Peak Signal-to-Noise Ratio ( $PSNR$ ) and Root Mean Square Error ( $RMSE$ ) to evaluate quality by calculating the error rate.

- 1) Mean Squared Error ( $MSE$ ).

$$MSE(x, y) = \frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2, \quad (20)$$

where  $N$  is number of pixels,  $x$  and  $y$  are signals, the error signal  $e_i = x_i - y_i$  is differences between two signal [29].

- 2) Peak Signal-to-Noise Ratio ( $PSNR$ ).

$$PSNR = 10 \cdot \log_{10} \left( \frac{L^2}{MSE} \right), \quad (21)$$

where  $L = 2^8 - 1 = 255$  for an 8-bit per pixel image [29]. High value of  $PSNR$  means better quality and less noise.

- 3) Root Mean Square Error ( $RMSE$ ).

$$RMSE(I, J) = \sqrt{MSE(I, J)} \\ = \sqrt{\frac{\sum_{j=1}^m \sum_{i=1}^n (I_{ij} - J_{ij})^2}{m \times n}}, \quad (22)$$

Where  $I, J$  are two image matrices [29].

Figures 5 and 6 show two images for one building from different angle, referred to as Data 0. Figures 7 and 8 show two images captured for same scene, referred to as Data 1. Figures 9, 10, 11 and 12 show the stitched images for the first, second, third and Fourth combinations of Data 1, respectively. Figures 13, 14, 15 and 16 show the stitched images for the first, second, third and Fourth combinations of Data 1, respectively. Figure 17 Provides analysis of the processing time in seconds as shown in (a), the  $PSNR$  as shown in (b) and  $RMSE$  as shown in (c), for the five combinations using 10 different



Fig. 5: First Image of Data 0



Fig. 6: Second Image of Data 0



Fig. 7: First Image of Data 1



Fig. 8: Second Image of Data 1

Two images of the same building were taken from two different angles as shown in Data 0 and Data 1.

scenes. The fifth combination refers to the method proposed in [2], as mentioned earlier. In Figure 17 (a), it is clear that the processing time of third and fourth combination take minimum time to process stitched panorama images. From Figure 17 (b) and Figure 17 (c) it is apparent that the fourth combination produce better result on most *PSNR* and *RMSE*. Table I shows the number of keypoints and matching points from testing Data 0. Table II shows the number of keypoints and matching points from testing Data 1. Table III show the first and second hybrid combinations results for 10 data (i.e. scenes) that contains two set of images with image resolution of  $1373 \times 2382$ . Table IV show the third and fourth hybrid combination results for 10 data. Table V show the fifth hybrid combination results for 10 data.

#### IV. DISCUSSION

This paper provided four different combinations that used for panorama stitching and compared their output images

based on processing time and quality. The third model reduced the processing time by 67% for the *ORB*, Brute Force (KNN), *RANSAC*, and Transformation Matrix compared to the fifth model, [2], that used *ORB*, Hamming distance, *PROSAC*, and the Transformation Matrix. The proposed fourth model reduced the processing time by 68% for the *ORB*, Brute Force (KNN), *PROSAC*, and Transformation Matrix compared to the fifth model, [2]. The fourth model that used *ORB*, Brute Force (KNN), *PROSAC*, and Transformation Matrix is shown better performance and quality results compared to other combinations. In particular, using *ORB*, Brute Force (KNN), *PROSAC*, and Transformation Matrix in the fourth model is shown better results of *PSNR* and *RMSE* compared to other combinations, as illustrated in Tables III, IV and V. Table I and Table II are showing the results for two different scenes that are referred to as Data 0 and Data 1. The four hybrid combinations are compared with regard to detector/descriptor type, where the *FAST* technique is used



Fig. 9: First Hybrid Combination



Fig. 10: Second Hybrid Combination



Fig. 11: Third Hybrid Combination



Fig. 12: Fourth Hybrid Combination

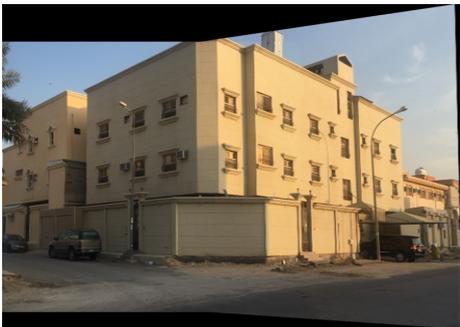


Fig. 13: First Hybrid Combination

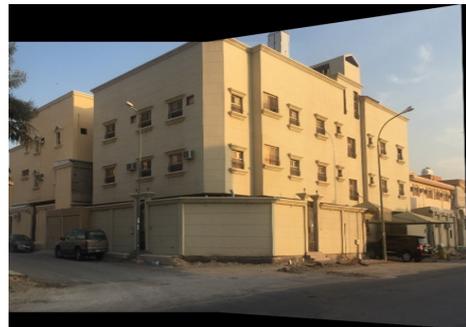


Fig. 14: Second Hybrid Combination



Fig. 15: Third Hybrid Combination



Fig. 16: Fourth Hybrid Combination

Figures 9, 10, 11 and 12 show the resulted four stitched images using Data 0, while Figures 13, 14, 15 and 16 show the resulted four stitched images using Data 1

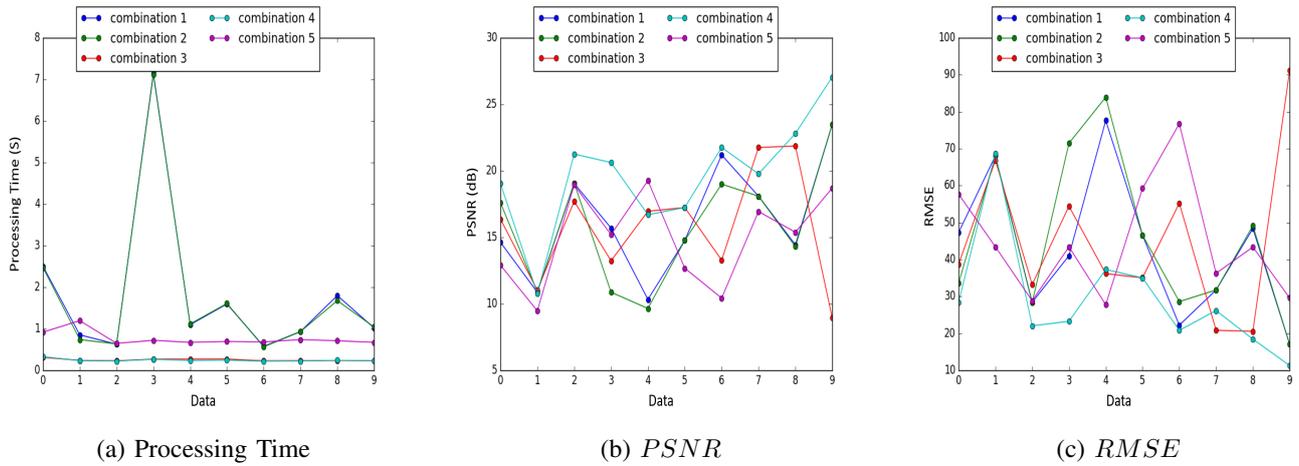


Fig. 17: Comparison of the five combinations in terms of: (a) Processing Time, (b) *PSNR* and (c) *RMSE* for 10 different data.

TABLE I: The Four Hybrid Combination Techniques Results of Data 0

Data 0								
	Detector/ Descriptor	Matching	Remove Outliers points	Alignment	Number of keypoints from first image	Number of keypoints from second image	Number of matched keypoints	Result
Method 1	<i>FAST</i> \ <i>BRISK</i>	<i>KNN</i>	<i>RANSAC</i>	Perspective Transform	7624	7731	31	Successful
Method 2	<i>FAST</i> \ <i>BRISK</i>	<i>KNN</i>	<i>PROSAC</i>	Perspective Transform	7624	7731	31	Successful
Method 3	<i>ORB</i> \ <i>ORB</i>	<i>KNN</i>	<i>RANSAC</i>	Perspective Transform	500	500	31	Successful
Method 4	<i>ORB</i> \ <i>ORB</i>	<i>KNN</i>	<i>PROSAC</i>	Perspective Transform	500	500	31	Successful

TABLE II: The Four Hybrid Combination Techniques Results of Data 1

Data 1								
	Detector/ Descriptor	Matching	Remove Outliers points	Alignment	Number of keypoints from first image	Number of keypoints from second image	Number of matched keypoints	Result
Method 1	<i>FAST</i> \ <i>BRISK</i>	<i>KNN</i>	<i>RANSAC</i>	Perspective Transform	2482	1901	31	Successful
Method 2	<i>FAST</i> \ <i>BRISK</i>	<i>KNN</i>	<i>RANSAC</i>	Perspective Transform	2482	1901	31	Successful
Method 3	<i>ORB</i> \ <i>ORB</i>	<i>KNN</i>	<i>RANSAC</i>	Perspective Transform	500	500	31	Successful
Method 4	<i>ORB</i> \ <i>ORB</i>	<i>KNN</i>	<i>PROSAC</i>	Perspective Transform	500	500	31	Successful

TABLE III: Comparison of the Four Hybrid Combination Techniques for 10 Data (1)

	Combination 1					Combination 2				
	Processing Time (s)	Contrast Differences		PSNR (dB)	RMSE	Processing Time (s)	Contrast Differences		PSNR (dB)	RMSE
		$d_A$	$d_B$				$d_A$	$d_B$		
Data 0	2.50309	365268	267576	14.631	47.312	2.46984	73218	112068	17.609	33.581
Data 1	0.85385	13023558	8068338	10.872	68.077	0.73972	13064142	8038308	10.872	68.077
Data 2	0.62836	85230	101964	19.053	28.437	0.63583	83136	91626	19.088	28.322
Data 3	7.14782	75186	87852	15.699	41.018	7.11124	1185936	1181022	10.875	71.48
Data 4	1.09822	1451064	1209600	10.332	77.618	1.11622	104508	115926	9.652	83.936
Data 5	1.59695	75132	72858	14.766	46.584	1.60874	99432	71598	14.766	46.586
Data 6	0.57794	85758	85872	21.194	22.224	0.56634	109380	102126	19.0	28.611
Data 7	0.93311	78144	93948	18.096	31.749	0.93277	80076	71814	18.096	31.749
Data 8	1.80077	122898	103104	14.439	48.371	1.67754	158718	122154	14.3	49.151
Data 9	1.02130	84066	84612	23.493	17.057	1.04310	84066	84516	23.492	17.058

TABLE IV: Comparison of the Four Hybrid Combination Techniques for 10 Data (2)

	Combination 3					Combination 4				
	Processing Time (s)	Contrast Differences		PSNR (dB)	RMSE	Processing Time (s)	Contrast Differences		PSNR (dB)	RMSE
		$d_A$	$d_B$				$d_A$	$d_B$		
Data 0	0.30996	80148	103404	16.368	38.739	0.32800	100956	100536	19.051	28.444
Data 1	0.24092	12035304	6983988	11.034	66.818	0.23476	12722154	8241258	10.788	68.739
Data 2	0.23189	124308	134700	17.702	33.222	0.22431	87174	92172	21.261	22.055
Data 3	0.26943	597486	576864	13.243	54.423	0.26545	105282	73854	20.623	23.27
Data 4	0.26986	104718	105714	16.961	36.182	0.23573	88644	98586	16.698	37.292
Data 5	0.27146	75126	85512	17.24	35.037	0.24401	77568	87516	17.243	35.028
Data 6	0.22970	1006734	801954	13.283	55.256	0.22012	83136	88596	21.759	20.825
Data 7	0.23260	65022	53256	21.752	20.843	0.22489	80082	69036	19.768	26.189
Data 8	0.23804	125682	110166	21.871	20.559	0.24186	108966	93636	22.808	18.456
Data 9	0.23218	315816	672126	8.934	91.165	0.23196	87876	89844	27.054	11.32

TABLE V: Comparison of the Four Hybrid Combination Techniques for 10 Data (3)

	Combination 5				
	Processing Time (s)	Contrast Differences		PSNR (dB)	RMSE
		$d_A$	$d_B$		
Data 0	0.91932	1203294	892212	12.906	57.705
Data 1	1.19511	49792386	35975232	9.475	43.431
Data 2	0.64943	944424	577704	18.934	28.829
Data 3	0.72190	1975944	1109010	15.203	43.431
Data 4	0.67391	552252	876252	19.258	27.774
Data 5	0.69564	1512900	1189476	12.681	59.225
Data 6	0.68438	1557420	729276	10.43	76.742
Data 7	0.73849	958362	1074060	16.943	36.258
Data 8	0.71170	2065578	1281894	15.384	43.385
Data 9	0.67448	2359956	832956	18.706	29.597

in first and second combination and the *ORB* technique is used in the third and fourth combinations. It can be seen that the FAST technique extracted more keypoints compared to the *ORB* technique. As minimum number of keypoints will reduce the processing time.

## V. CONCLUSION AND FUTURE WORKS

Researchers have worked to improve panorama stitching techniques and minimize the computational requirements. This paper focused on a new application of panorama stitching for a 2D scenario. The proposed methods can generate a semi-transparent view as a solution for the project enabling drivers to effectively see around corners. This paper also compared the quality and processing time of the produced 2D views within the scope of state-of-the-art methods. For future work, the presented models can be extended to include 3D scenes. Another future work can provides a real-time processing for similar applications to assist drivers and pedestrians. Finally, additional novel techniques could be implemented to enhance the methods discussed in this work.

## ACKNOWLEDGMENT

We would like to express our special thanks to King Faisal University and to Dr. Syed Afaq Husain, Dr. Muhammad Bilal Ahmad and Dr. Asrar Ul Haque for their feedback which helped to improve the project.

## REFERENCES

- [1] Haque, M.J., "Improved Automatic Panoramic Image Stitching," *Lap Lambert Academic Publishing GmbH KG*, 2012
- [2] H. Jeon and J. Jeong and K. Lee, "An implementation of the real-time panoramic image stitching using *ORB* and *PROSAC*," *International SoC Design Conference (ISOCC)*, 2015, pp. 91–92.
- [3] M. Wang and S. Niu and X. Yang, "A novel panoramic image stitching algorithm based on *ORB*," *International Conference on Applied System Innovation (ICASI)*, 2017, pp. 818–821.
- [4] A. Agarwal and M. Agrawala and M. Cohen and D. Salesin and R. Szeliski, "Photographing long scenes with multi-viewpoint panoramas," *ACM TRANSACTIONS on Graphics*, 2006, vol. 25, pp. 853–861.
- [5] R. Gupta and R. I. Hartley, "Linear pushbroom cameras," *IEEE TRANSACTIONS on Pattern Analysis and Machine Intelligence*, 1997, vol. 19, pp. 963–975.
- [6] S. Peleg and B. Rousso and A. Rav-Acha and A. Zomet, "Mosaicing on adaptive manifolds," *IEEE TRANSACTIONS on Pattern Analysis and Machine Intelligence*, 2000, vol. 22, pp. 1144–1154.
- [7] A. Zomet and D. Feldman and S. Peleg and D. Weinshall, "Mosaicing new views: the Crossed-Slits projection," *IEEE TRANSACTIONS on Pattern Analysis and Machine Intelligence*, 2003, vol. 25, pp. 741–754.
- [8] M. Agrawala and D. Zorin and T. Munzner, "Artistic Multiprojection Rendering," *Proceedings of the Eurographics Workshop on Rendering Techniques*, 2000, pp. 125–136.
- [9] J. Yu and L. Mcmillan, "A Framework for Multiperspective Rendering," *Proceedings of the 15th Eurographics Conference on Rendering Techniques*, 2004, pp. 61–68.
- [10] J. Yu and L. Mcmillan, "General Linear Cameras," *Proceedings of the 8th European Conference on Computer Vision*, 2004, pp. 14–27.
- [11] M. Z. Bonny and M. S. Uddin, "Feature-based image stitching algorithms," *International Workshop on Computational Intelligence (IWCI)*, 2016, pp. 198–203.
- [12] R. Szeliski and H. Shum, "Creating Full View Panoramic Image Mosaics and Environment Maps," *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques [ACM Press]*, 1997, pp. 251–258.
- [13] A. Wójcicka and Z. Wróbel, "The Panoramic Visualization of Metallic Materials in Macro- and Microstructure of Surface Analysis Using Microsoft Image Composite Editor (ICE)," *Proceedings of the Third International Conference on Information Technologies in Biomedicine*, 2012, pp. 358–368.
- [14] P. Azad and T. Asfour and R. Dillmann, "Combining Harris interest points and the SIFT descriptor for fast scale-invariant object recognition," *International Conference on Intelligent Robots and Systems*, 2009, pp. 4275–4280.
- [15] J. Jiao and B. Zhao and S. Wu, "A speed-up and robust image registration algorithm based on FAST," *IEEE International Conference on Computer Science and Automation Engineering*, 2011, vol. 4, pp. 160–164.
- [16] L. Yu and Z. Yu and Y. Gong, "An Improved *ORB* Algorithm of Extracting and Matching Features," *International Journal of Signal Processing and Pattern Recognition*, 2015, vol. 8, pp. 117–126.
- [17] E. Rosten and T. Drummond, "Machine Learning for High-speed Corner Detection," *Proceedings of the 9th European Conference on Computer Vision - Volume Part I [Springer-Verlag]*, 2006, pp. 430–443.
- [18] J.J. Anitha and S.M. Deepa, "Tracking and Recognition of Objects using SURF Descriptor and Harris Corner Detection," *International Journal of Current Engineering and Technology*, 2014, vol. 4, pp. 775–778.
- [19] K. Dohi and Y. Yorita and Y. Shibata and K. Oguri, "Pattern Compression of FAST Corner Detection for Efficient Hardware Implementation," *21st International Conference on Field Programmable Logic and Applications*, 2011, pp. 478–481.
- [20] E. Rublee and V. Rabaud and K. Konolige and G. Bradski, "*ORB*: An efficient alternative to SIFT or SURF," *International Conference on Computer Vision*, 2011, pp. 2564–2571.
- [21] M. Brown and D.G. Lowe, "Automatic Panoramic Image Stitching using Invariant Features," *International Journal of Computer Vision*, 2007, vol. 74, pp. 59–73.
- [22] C. Harris and M. Stephens, "A Combined Corner and Edge Detector," *Proceedings of the 4th Alvey Vision Conference*, 1988, pp. 147–151.
- [23] S. Leutenegger and M. Chli and R. Y. Siegwart, "*BRISK*: Binary Robust invariant scalable keypoints," *International Conference on Computer Vision*, 2011, pp. 2548–2555.
- [24] A. S. Arefin and C. Riveros and R. Berretta and P. Moscato, "GPU-FS-KNN: A Software Tool for Fast and Scalable KNN Computation Using GPUs," *PloS one*, 2012, vol. 7, pp. e44000.
- [25] J. S. Beis and D. G. Lowe, "Shape indexing using approximate nearest-neighbour search in high-dimensional spaces," *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1997, pp. 1000–1006.
- [26] O. Chum and J. Matas, "Matching with *PROSAC* - progressive sample consensus," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005, vol. 1, pp. 220–226.
- [27] P. Ostiak, "Implementation of HDR panorama stitching algorithm," *10th Central European Seminar on Computer Graphics for Students (CESCG)*, 2006.
- [28] C.Y. Chen, "Image Stitching - Comparisons and New Techniques," *CITR-TR-30*, 1998.
- [29] P. Ndajah and H. Kikuchi and M. Yukawa and H. Watanabe and S. Muramatsu, "An investigation on the quality of denoised images," *International Journal of Circuits, Systems and Signal Processing*, 2011, vol. 5, pp. 423–434.



# A Social Bonds Integration Approach for Crowd Panic Simulation

Imene Bouderbal

Ecole Nationale Preparatoire aux Etudes  
d'Ingeniorat, Rouiba, Algiers, Algeria  
Email: imene.bouderbal@yahoo.com

Abdenour Amamra

Ecole Militaire Polytechnique,  
Bordj El-Bahri BP 17, Algiers, Algeria  
Email: amamra.abdenour@gmail.com

**Abstract**—Crowd panic has incurred massive injuries and deaths throughout history; thus understanding it is particularly important in order to save human lives. Recently, numerous simulation methods have been contributed in order to provide insight into the design of evacuation planning strategies. In this paper, we integrate a social structure to the crowd mobility model for the purpose of investigating the influence of social bonds on collective behavior during panic. A macroscopic crowd panic model based on social science theories was integrated as an internal module to the microscopic mobility model. The resulting framework is tunable and permits the implementation of several panic scenarios. It is also designed to run in different situations for a better comprehension of panic-related phenomena. The results demonstrate the smoothness of our crowd flow model and the realism of evacuation during panic.

**Index Terms**—Panic modeling, Crowd simulation, Human behavior in panic, Evacuation disaster.

## I. INTRODUCTION

THROUGHOUT history, mankind has experienced several disasters and accidents caused by panic in overcrowded spaces (Fig 1a). Hence, simulating crowd with panic has an important value in emergency planning for architectures characterized by a high-density crowd. Special attention should be dedicated to a deeper understanding of the nature of panic during the process of escaping hazardous situations within buildings and public spaces [1]. On the other hand, a clearer understanding of the social interactions in difficult situations would give more insights on how we could prevent casualties beforehand by adapting the design of buildings to the potential course of events during an evacuation situation. The understanding of such behavior brings a significant architectural added value, which in turn would result in safer and more reliable buildings.

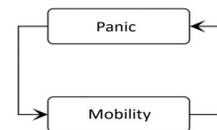
The prediction of the birth of collective panic and the possible actions that might cause material damage and human losses can be difficult. Modeling and simulation constitute a solution to tackle such a problem. These tools allow us to generate virtual environments including both human agents and buildings, an evacuation scenario can then be run under some assumptions on a given emergency scheme.

Our objective is the modeling of panic propagation process and its consequences on human behavior while taking into account the social aspect of the crowd. To this end, the remainder of the paper is structured as follows: in Section

2 we describe the main related works that were contributed in the field of crowd simulation. in Section 3 we present a brief description of crowd behavior during panic situations. Section 4 is the core of our work, where we propose a socially-aware panic model. In Section 5, we present the results of our contribution and we discuss their impact compared to alternative methods. Finally, we conclude the paper, and we trace the potential perspectives that can extend the current work.



(a)



(b)

Fig. 1: (a) Crowd in panic, (b) Systemic loop of panic.

## II. RELATED WORKS

Crowd simulation models have roots in different domains such as entertainment, security, urbanism, architecture, management, etc. Since we are interested in panic modeling, we will present the main works that have permitted the modeling of panic evolution phenomenon and its effects on the crowd as well as those which gave consistent results with empirical observations. Indeed, we will focus on conceptual modeling rather than the available software tools. After analyzing collective behaviors literature, three main theoretical approaches were found.

The first one, called *Contagion Theory* [2, 3], it states that an individual in a crowd loses their conscious personality and obeys to all suggestions from the crowd influent members. Works belonging to this category incorporate the concept of social forces, where the model allows a mixture of socio-psychological and physical forces influencing the behaviors within a crowd [4]. Several models based on this approach have been developed so far, the latter proposes an adaptation and hybridization of the model of social forces with other models [5]. These models can reproduce certain phenomena observed in real crowds during panic movements, such as crowd oscillation in a narrow passage, arch formation in front

of an exit, lane formation, etc. However, they neglect the description of panic evolution phenomenon and reproduce only some of its aspects. Moreover, the social structure of the crowd is neglected and the evolution of the emotional state of individuals is not taken into account, which impinges on the realism.

The second approach is based on the *Emergent Norm Theory* [6], which considers that the collective behavior is an outcome of the interactions among individuals, which are able to evaluate the information they receive and to decide on its usage at the present situation. In fact, agents assign positive or negative values to the information, which leads to the development of an interactive cognition. This approach analyzes the agent's micro-properties that help in the social system formation. It also analyzes the behavior patterns at a group level. An example of works based on this approach can be found in [7], where the authors describe an individual behavior model for agents. Although there is no explicit way of interaction or communication, each agent is aware of itself and their peers.

The last approach is a variant of the second one, commonly called *Structuralism*, it inverts the formula and emphasizes the social structure studies and their impact on individuals. It is a macro-to-micro approach since it considers that the changes are triggered by the crowd (macro) to the individuals (micro). Social science research such as [8] embraced the structural theory approach. One of the works that had implemented the macro-to-micro approach for panic crowd situations is [9], where the authors utilized system dynamics and consider that panic is a domino effect. In fact, the movements of individuals result from two complementary models that interact and influence each other mutually, namely the mobility model and the panic model (Fig 1b).

These models are more informative than those based on alternative approaches. In fact, they allow us to describe the beginning and the evolution of panic, thereby to adapt the behaviors to the individual state. However, the solutions that were found model sudden transitions between different states of human emotions. In fact, before an individual behaves in a non-rational fashion, they go through a stage called *Limited Rationality* and can remain in it or calm down without reaching the state of panic. Also, these models do not take into account the social structure of the crowd.

### III. CROWD PANIC

Understanding how crowds behave during critical situations has long been necessary for emergency response and management. In fact, most of the normal behavior vanishes when pedestrians face an emergency situation (it does not always have to be an emergency situation, however, similar actions can be observed for example among crowds trying to get the best seats at a concert or consumers running for sales) and non-adaptive crowd behaviors appear. These behaviors are recognized to be responsible for the death and injury of most victims in crowd disasters. Non-adaptive crowd behaviors refer to the destructive actions that a crowd may experience

in emergency situations, such as stampede, pushing, knocking and trampling on others [1]. Generally, these behaviors are the result of a dangerous psycho-sociological phenomenon, which is *panic*.

Panic is a phenomenon generally studied in psychology and human science and often identified by its consequences. It is triggered whenever a situation of tension worsens, slips or escapes from human control. Panic is defined as an intense fear triggered by the occurrence of a real or imaginary danger felt simultaneously by all individuals in a group, a crowd or a population, characterized by the regression of mentalities to an archaic and gregarious level, leading to primitive reactions of hopeless jumps, indiscriminate agitation of violence or collective suicide [5]. Nevertheless, discrepancies exist between the definition of triggering and propagation processes. This difference gives rise to different social theories [1]. The latter differs in the definition of the relationship between disasters and panic phenomenon, and the type of behavior in a state of panic. Despite the differences, these approaches come together on two important points:

- Panic is a feeling of extreme fear that invades the pedestrian following the perception of possible danger;
- Social interaction in a crowd promotes the spread of panic.

### IV. PROPOSED MODEL

In order to ensure the safety of people and reduce the impact of panic caused by disasters, it is necessary to understand and model this dangerous phenomenon, its propagation process, and its consequences. Panic model presented hereinafter is based on system dynamics. It has been integrated into a mobility model for crowd evacuation simulation during catastrophes. Besides, it uses the same environment and pedestrian modeling (Fig 1b). The proposed mobility model is microscopic, whereas the proposed panic model is macroscopic (Fig 2). Thereby, panic model permits the study of the evolution of panic phenomenon at a different level of granularity than that adopted for the study of mobility. Many types of relations coexist between the individuals of a crowd (family bonds, friendships, professional relations, etc.). Behaviors and decisions made by these individuals depend strongly on the social relations [10, 11]. We added a social model that describes the social structure of the crowd and its evolution during the emergency in order to study their effects on pedestrians mobility and the propagation of panic.

#### A. Social structure of the crowd

In order to build an innovative solution for crowd modeling, we chose to integrate an important aspect into our mobility model which is the *social structure of the crowd*. This latter could be integrated by using a social model [2]. The one we adopted in our work is *Small World* because its generated graphs are the closest to the real representation of contacts between individuals and groups and their evolution in space and time. These graphs allow us to predict the evolution of the epidemic and by analogy the evolution of panic.

We used a new model for generating small world graphs. Transformations and constructions that it orchestrates follow sociological behaviors observed and described in the literature. The generation process provided by this model is divided into three phases, each permits to accomplish a different aspect of links evolution in the crowd.

1) *Initialization*: The model we used is configurable. The generative process behind it permits to create a simple, connected and undirected graph  $G = (V; E)$ , of  $|V|$  nodes and  $|E|$  edges.

2) *Creating nodes (inserting agents into the simulated environment)*: The first step allows the addition of new individuals in the social structure of the crowd. In fact, every time an agent is added to the simulation environment, they join the social structure. The graph thus obtained contains the expected number of nodes but does not reach the expected number of edges. This first step returns a connected simple, acyclic graph, resulting in a topology similar to that of a tree. It represents the initial social links that will evolve during the simulation.

3) *Adding random edges (connecting agents)*: The second phase allows us to link members of the crowd randomly. This step serves to reproduce the process of connecting people that know each other, either before joining the network or afterward. In order to implement different types of social links, we included a weighting strategy to the links. Each edge will be assigned a value that represents its type. We model three types of links: family, friendship link and professional.

4) *Reinforcement of communities*: This stage corresponds to the moment when the members of the network begin to establish relationships with other members who share the same friends. Two persons who didn't know each other beforehand can potentially become friends, which will lead to the creation of new links between them. The creation of edges according to this scheme allows the strengthening of communities, thereby increasing the agglomeration coefficient of the generated network. Unlike the first two phases that run before the launch of the simulation, the latter runs simultaneously with the mobility model.

### B. Adaptation of Susceptible, Infectious and Recovered (SIR) epidemic model for panic propagation

The panic model was integrated into the mobility model and it uses the same environment and pedestrian modeling. The analogy between panic and contagion led us to a propagation model inspired by the epidemiological ones, and strengthened by the work of [12] and [3]. We adapted the SIR compartmental model [13] to describe the evolution and spread of panic within a crowd. Compartmental models are a technique used to simplify the mathematical modeling of infectious disease. For instance, the population is divided into compartments, with the assumption that all the individual within the same compartment share the same properties [13]. These models are usually modeled through ordinary differential equations (which are deterministic), but can also be viewed in a stochastic framework, which is as realistic as complicated to analyze (Fig 2a). Our version of the SIR model consists of four

compartments where individuals are classified according to their emotional state:

- SP: compartment of people susceptible to panic.
- RL: compartment of people who behave with limited rationality. The individuals of this compartment have not yet reached the state of panic but are already disturbed.
- NR (P): compartment of people in a state of panic. The individuals of this compartment have reached the state of panic and behave irrationally.
- NP: compartment containing non-panicked people (calm).

In fact, an individual is not assigned to the compartment NP until he reaches the emergency exit and leaves the simulation environment. Furthermore, we added an intermediate step before the transition to the state of panic, which is modeled via the RL compartment. It was introduced to express the non-spontaneous aspect of panic. This latter appears when the negative emotions of the individual (fear, stress) reach a predefined threshold. This last can be modeled by latency periods in the RL compartment that are randomly chosen to be able to express the heterogeneity among individuals. We also admit that some individuals have personal predispositions to panic. They will be in a panic state at the beginning of the simulation and will represent the initial panic transmission vector.

#### 1) Transitions between compartments:

- Each susceptible individual moves from the compartment SP to the compartment RL with a transition rate ( $\alpha_1$ ) following direct contact with a panicked person.
- ( $\alpha_2$ ) represents the transition rate from the compartment RL to the compartment SP. This transition expresses the possibility that an individual RL can calm down after a certain period of time without reaching the state of panic.
- ( $\beta_1$ ) is the transition rate for which RL individuals transit to the compartment NR (P). In fact, each individual of the RL compartment spends some time in the latter. The choice of this parameter is very important because it allows changing phases in the process of panic propagation between an active phase, when panic spreads, and an inactive phase when it disappears.
- ( $\beta_2$ ) is the transition rate for which panicked people pass to the RL compartment. This transition happens for each individual after being panicked for a certain period of time. This aspect has been added to our model to take into account the ephemeral aspect of panic.

### C. Integration of the intermediate models

In order to test and validate the proposed solutions, we integrated them in our mobility model and we adapt this latter to the modifications added. We start by integrating the social structure of the crowd obtained via the social model Small World in the decision making the process of agents. Afterward, we integrate panic model developed as an internal module in the mobility model.

1) *Integration of the social structure into the decision-making process*: Individuals will, now, adjust their displacement vector so as to evacuate while joining members that

belong to their social group. This strategy was implemented using the distance maps obtained by the *FMM* method and provided by one of the two upper layers of the environment model in Fig 2b. Hence, each agent will have the possibility to identify the positions of individuals with whom they are socially linked and to choose the member to join according to the type of the link. The decision-making process consists of the following iterative steps:

- 1) Collect distance maps provided by the environment layers;
- 2) Collect the map of social links;
- 3) Identify the member to join among those having a social link with it;
- 4) Trigger a movement decision based on the collected data;
- 5) Perceive near environment information;
- 6) Adjust the decision according to the social forces model and behavior rules;
- 7) Take into consideration the individual properties that influence the decision;
- 8) Execute the decision.

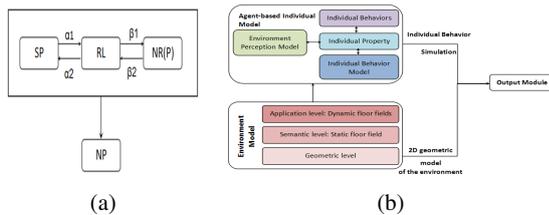


Fig. 2: (a) Proposed panic model, (b) Mobility model architecture.

2) *Integration of panic model*: The integration of panic model will change the system's flow. In fact, mobility and panic interact and influence each other mutually. Panic is nothing more than a structure contained in the crowd that will be activated when certain conditions are met. The general process of unfolding panic model follows two main phases. The first one takes place before the launch of the simulation. It allows the user to initialize the individuals initially panicked. Whereas, the second takes place during the simulation. In fact, it runs in parallel with the mobility model.

3) *Panic consequences modeling*: Once an individual becomes panicked, he behaves in a non-rational manner and adopts a social attitude toward the rest of the crowd, even toward members of their own social group. This behavioral disorder is reflected in our mobility model by the disappearance of social values and the emergence of violence and individualism. Violence was modeled by incrementing repulsive forces and individualism through the non-consideration of the social structure in the decision-making process. Nevertheless, panicked behaviors vary according to the nature of each person. We modeled this heterogeneity through different criteria and individual characteristics such as aggressiveness, restriction of the field of vision, increasing the speed of

movement or reducing the comfort distance.

## V. RESULTS AND DISCUSSION

In this section, we study the flow while taking into account the social structure of the crowd and panic phenomenon. The validation of the mobility model was carried out on two steps: qualitative and quantitative. Several qualitative aspects of pedestrian dynamics were reproduced by our model such as lane formation, the arch phenomenon, oscillations at the bottleneck, etc (Fig 3a), and no more unrealistic congestions are encountered (Fig 3b). The quantitative study that we conducted regards pedestrians' flow through a narrow path (bottleneck) (Fig 3c). The experimental setup that we simulated was inspired by [14]. The results obtained using our mobility model correspond well to the real-world observations (Tab I, II).

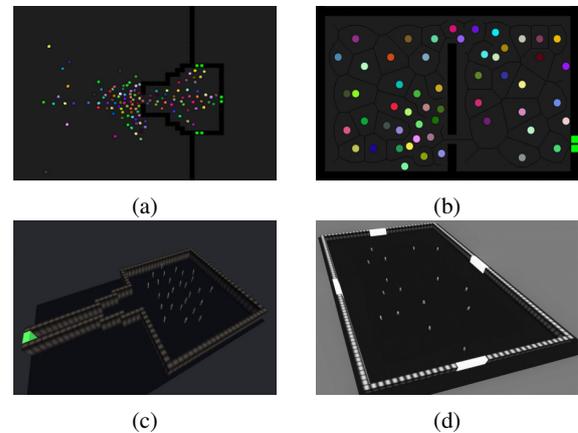


Fig. 3: (a) (b) Qualitative aspects obtained by the mobility model, (c) Bottleneck generation with our simulation framework, (d) Real simulation scenario

### A. Study of social structure impact

First, we start with the simulation results obtained without introducing the social structure into the mobility model in

TABLE I: Experimental specific flow  $J_{s,exp}$  [14]

b(m)	$N_b=20$	$N_b=40$	$N_b=60$
0.8	1.86	1.77	1.61
0.9	2.06	1.91	1.86
1.0	2.19	2.08	1.9
1.1	1.78	1.93	1.93
1.2	2.31	1.81	1.97

TABLE II: Values of the specific flow  $J_s$  obtained using our Framework.

b(m)	$N_b=20$	$N_b=40$	$N_b=60$
0.8	1.73	1.45	1.40
0.9	1.92	1.49	1.53
1.0	2.07	1.63	1.66
1.1	2.26	1.78	1.73
1.2	2.35	1.83	1.86

order to study the impact of this latter on the decision-making process without introducing panic. In order to distinguish social links influence, we run a series of simulations using the same scenario with the new and the old version of the mobility model. The test scenario we used is inspired by that illustrated in (Fig 3d).

Several qualitative aspects of pedestrian dynamics were reproduced by our model. Many of which were already obtained without the social aspect of the crowd such as, lane formation, the arch phenomenon, oscillations at the bottleneck. However, by introducing the social aspect congestion reappears, as well as the formation of reconciliation movements and training groups of different sizes, which is closer to reality.

### B. Study of panic propagation and its effects

In order to simplify the study of panic propagation, we simulate without mobility. Since there is not a method to quantify the growth rates of emotion intensity [3], we vary the latency and infection period from one individual to another to express their emotional and psychological heterogeneity. Nevertheless, the proposed panic model is macroscopic, it describes the evolution of the population and not that of the individuals themselves. This evolution is considered continuous over time, which is reflected mathematically by an ordinary differential equation system :

$$\begin{cases} \frac{dx}{dt} = -\alpha_1 \cdot x + \alpha_2 \cdot y \\ \frac{dy}{dt} = \alpha_1 \cdot x - \left(\frac{\beta_1}{N} \cdot z + \alpha_2\right) \cdot y + \beta_2 \cdot z \\ \frac{dz}{dt} = \left(\frac{\beta_1}{N} \cdot y - \beta_2\right) \cdot z \end{cases} \quad (1)$$

The ratio  $\frac{\beta_1}{N \cdot \beta_2} \cdot y$  defines the basic reproduction number  $R_0$ . According to its value, panic will spread or shrink. Generally, it is the initial value  $R_0$  that is used. We have either:

- $R_0 = \frac{\beta_1}{\beta_2} > 1$  : propagation,
- $R_0 = \frac{\beta_1}{\beta_2} < 1$  : shrinkage.

Several methods exist to solve the differential equations system. We used the approximation obtained by the classical *Runge-Kutta* and we implemented it then integrated it to our simulation platform. In order to validate the hypothesis about panic propagation, we conducted an experiment inspired by the one presented in [3]. We start with a population of 50 individuals susceptible to panic, 70 individuals of limited rationality and 5 panicked individuals. We vary the value of the transition rates  $\beta_1$  and  $\beta_2$  to verify the validity of the expression of the basic reproduction number. Fig 4a and 4b show the evolution of the numbers of individuals in each compartment for  $R_0$ , respectively, greater than and less than 1.0.

The obtained results show that the spread of panic and its shrinkage depend on the choice of the user. Moreover, the behavior of individuals can be defined according to their state. This aspect allows for implementing several cases in order to study and evaluate faithfully safety procedures. By introducing mobility, the social structure of the crowd evolves

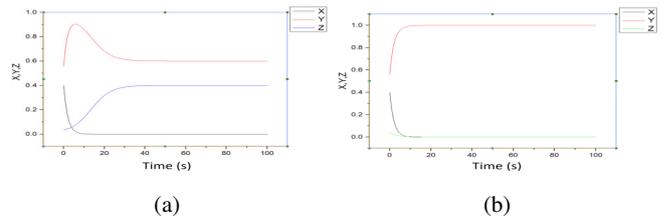


Fig. 4: Graphs showing the evolution of the number of SP, RL and P individuals for a basic reproduction number (a)  $R_0 > 1$  and (b)  $R_0 < 1$ .

and promotes the spread of panic and allows it to reach isolated groups.

In order to position our model among the existing ones, we conducted a comparison between the models [3, 15, 16] and the proposed one according to the following criteria: reactive/cognitive approach, micro-level explicitness, macro-level explicitness, communication forms, and panic behavior explicitness.

The reactive and cognitive approaches are related to how individuals are modeled in the system. Authors in [3, 15] adopted both the cognitive approach, while [16] and the proposed model combine both approaches.

The micro and macro levels parameters refer to the presence of components that represent the collectivity in an explicit fashion. Concerning the micro level explicitness: [16] and [3] deal with it using agent-based models, [15] on the other hand, uses collective agents and our model uses a hybridization of individual behavior model, social forces model, and global rules. The macro level explicitness, however, was addressed using: group formation with social forces by [16], imaginary and group mind formation by [15], without detail in [3] and using a Social crowd model in our solution.

The communication form parameter deals with the interactions among the collectivity of agents. These interactions can happen either directly, indirectly or through the process of perturbation and dissipation. The communication was of this latter type in [15] and, indirect in [16], and [3] as well as in the proposed model.

Finally, panic behavior parameter deals with the usage of the collective behavior formation stage. If such a stage is modeled and the transition is described in detail for each agent, they could behave more realistically, and the simulation could be closer to reality. This aspect was not available in [16], and was modeled using: a framework based on symbolic interactionism in [15], a macroscopic model for panic spread based on an epidemiological approach in [3] and in the proposed solution.

## VI. CONCLUSION AND PERSPECTIVES

In this paper, we developed a macroscopic panic model that permits to study the evolution of a panicked population. This model is based on an epidemiological approach and takes into account the social aspect of the crowd. We also integrated this model as an internal component of a microscopic mobility model to investigate and explore the complementarity between

macroscopic and microscopic models and the possibility of enrichment between them. The obtained results are encouraging and replicate faithfully the real crowd evacuation situations.

As further work, we think there are topics that need to be deeply investigated in order to improve crowd simulation and analysis. First, human emotions represent one of the most important factors that affect the propagation of panic in the crowd. Thus, modeling emotions and their evolution process will systematically enhance the realism of panic propagation. Second, the role of an individual during the emergency situation is also one of the most important factors, which affect the process of decision making. Therefore, modeling the different roles that can be assigned to individuals and computing their influence on the crowd will be necessary for an even more realistic simulation. Third, human behavior depends on the state of each individual in the crowd, the latter can be obtained from sociological studies and empirical observations.

#### REFERENCES

- [1] Anthony R Mawson. "Understanding mass panic and other collective responses to threat and disaster". In: *Psychiatry: Interpersonal and biological processes* 68.2 (2005), pp. 95–113.
- [2] Tibor Bosse et al. "Modelling collective decision making in groups and crowds: Integrating social contagion and interacting emotions, beliefs and intentions". In: *AUTON AGENT MULTI-AG* 27.1 (2013), pp. 52–84.
- [3] Haifa Abdelhak. "Modélisation des phénomènes de panique dans le cadre de la gestion de crise". PhD thesis. Université du Havre, 2013.
- [4] Zhilu Yuan et al. "Simulation model of self-organizing pedestrian movement considering following behavior". In: *Frontiers of Information Technology & Electronic Engineering* 18.8 (2017), pp. 1142–1150.
- [5] Peng Lin, Jian Ma, and Siuming Lo. "Discrete element crowd model for pedestrian evacuation through an exit". In: *Chinese Physics B* 25.3 (2016), p. 034501.
- [6] Mei Ling Chu et al. "Modeling social behaviors in an evacuation simulator". In: *C. Animation and Virtual Worlds* 25.3-4 (2014), pp. 373–382.
- [7] Jinhuan Wang et al. "Modeling and simulating for congestion pedestrian evacuation with panic". In: *Physica A: Statistical Mechanics and its Applications* 428 (2015), pp. 396–409.
- [8] Russell Hardin. *Collective action*. RFF Press, 2015.
- [9] Lu Tan, Mingyuan Hu, and Hui Lin. "Agent-based simulation of building evacuation: Combining human behavior with predictable spatial accessibility in a fire emergency". In: *Inf. Sciences* 295 (2015), pp. 53–66.
- [10] Yan Li et al. "A grouping method based on grid density and relationship for crowd evacuation simulation". In: *Physica A Stat. Mech. Appl.* 473 (2017), pp. 319–336.
- [11] Dirk Helbing and Peter Molnar. "Social force model for pedestrian dynamics". In: *Physical review E* 51.5 (1995), p. 4282.
- [12] Farid Kadri, Babiga Birregah, and Eric Châtelet. "The impact of natural disasters on critical infrastructures: A domino effect-based study". In: *J. of Homeland Security and Emergency Management* 11.2 (2014), pp. 217–241.
- [13] Fatima Zohra Younsi. "Mise ne palce d'un système d'aide à la décision pour le suivi et la prévention des épidémies". PhD thesis. Université d'Oran, 2016.
- [14] Armin Seyfried et al. "New insights into pedestrian flow through bottlenecks". In: *Transportation Science* 43.3 (2009), pp. 395–406.
- [15] Robson dos Santos França, Maria das Graças Bruno Marietto, and Margarethe Born Steinberger. "A Multi-agent Model for Panic Behavior in Crowds". In: (2009).
- [16] Bachar Kabalan. "Crowd dynamics: modeling pedestrian movement and associated generated forces". PhD thesis. Université Paris-Est, 2016.

# GNSS-based Sound Card Synchronization

Alexander Carôt  
Anhalt, University of Applied Sciences  
Lohmannstr. 23  
06366 Köthen, Germany  
Email: alexander.carot@hs-anhalt.de

Hasan Mahmood  
Symonics GmbH  
Geierweg 25  
72144 Dußlingen, Germany  
Email: hasan.mahmood@symonics.com

Christian Hoene  
Symonics GmbH  
Geierweg 25  
72144 Dußlingen, Germany  
Email: christian.hoene@symonics.com

**Abstract**—Audio communication on the public Internet suffers from not synchronized word clocks of the involved audio devices. The resulting clock drift leads to audio dropouts, which is typically compensated by a sample rate conversion (SRC) in standard telecommunication systems. This, however, does not fulfill the requirements of a high-quality audio system, in which all devices share one and the same word clock. Professional IP based network audio systems such as DANTE or AVB with their respective clock synchronization techniques have so been limited to LAN usage, where network jitter and loss have negligible importance regarding the required accuracy in the dimension of several nanoseconds. In a WAN, however, jitter in the millisecond dimension would lead to unacceptable measurement errors for the intended clock synchronization. As a consequence, we decided to investigate alternative clock synchronization techniques for WAN-distributed devices and developed a GNSS-based approach, which leads to precise clock synchronization.

## I. INTRODUCTION AND PROBLEM

The term "distributed music" or "network music performance" describes a scenario, in which at least two dislocated musicians perform together as if being in the same room. This domain has been investigated for more than two decades [3]. However, with the increasing stability of nowadays available broadband networks another quality reducing factor has become relevant: Despite commonly applied standard audio sample rates of 44,1 kHz, 48 kHz or 96 kHz [10] the word clocks of two different devices do not run in precise synchrony for physical reasons. With respect to audio networking, clock drift means that one audio process is running faster than the remote one. As a consequence, the faster process will not receive a sufficient amount of audio samples, which in turn leads to a buffer underrun and a corresponding audio dropout in specific intervals ranging between 10 and 30 seconds depending on the actual amount of drift. On the other side, the slower process receives too many samples, which eventually leads to a buffer overrun in the same interval, which also corresponds to disturbances in the audio signal. In context with low-latency audio networking the network buffers should be adjusted as low as possible, however, due to the described clock drift problem this proportionally increases the probability for audio dropouts.

In LAN-based sound systems (Local Area Network) such as Dante [2] or AVB [7], each device of the audio network

fast-music is part of the fast-project cluster (fast actuators sensors & transceivers), which is funded by the BMBF (Bundesministerium für Bildung und Forschung).

is therefore synchronized to either a dedicated master clock or a specific device on the network, which was previously identified as the clock master. With respect to our Internet-based application, this synchronization process typically cannot be applied due to the existing network jitter in the common dimension of at least one millisecond and rather more. In [4] we presented an approach, which is able to provide WAN-based synchronization (Wide Area Network) by averaging the resulting measurement error, however, the reliability of this approach is limited depending on the actual amount of network jitter.

## II. CONCEPT

In this section, we present a novel concept, which eventually provides reliable sound card synchronization for WAN-distributed devices. Our previous and not perfectly reliable approach uses the WAN itself as the source of synchronization. In contrast, our new concept takes advantage of global navigation satellite systems (GNSS) as the synchronization link between the involved sound devices. We consider GNSS such as the global positioning system (GPS) excellent sources for a grandmaster clock in order to synchronize the word clocks of the involved devices. GNSS transceivers provide a 1-PPS (one pulse per second) output signal, which gives a high pulse at every start of a second [5]. This pulse is our reference point to the absolute start of a second. The UTC (Coordinated Universal Time) time received from GNSS can be used to synchronize the word clocks, but there is a delay between the time received by the GNSS module and time set in the processor, hence it is not accurate up to microsecond level. In order to compensate this offset we take advantage of the time stamping capabilities of IEEE 1588 clocks, which capture the moment of the actual pulse occurrence so that the local time can be set later according to the precise reference. Despite the synchronization with the 1-PPS pulse the local oscillator and the GNSS clock still suffer from the inherent clock drift so that the clocks go out of sync after a couple of seconds. We overcome this problem via a feedback loop and a proportional, derivative and integral (PID) controller that keeps track of the clock drift and adjusts the clock rates respectively. The 1588 clock follows the GNSS clock to a precision and accuracy of greater than one microsecond. Once the 1588 clock is synchronized, we apply the appropriate fine-tuning to an audio clock PLL (phase locked loop) accordingly. Figure 1 illustrates

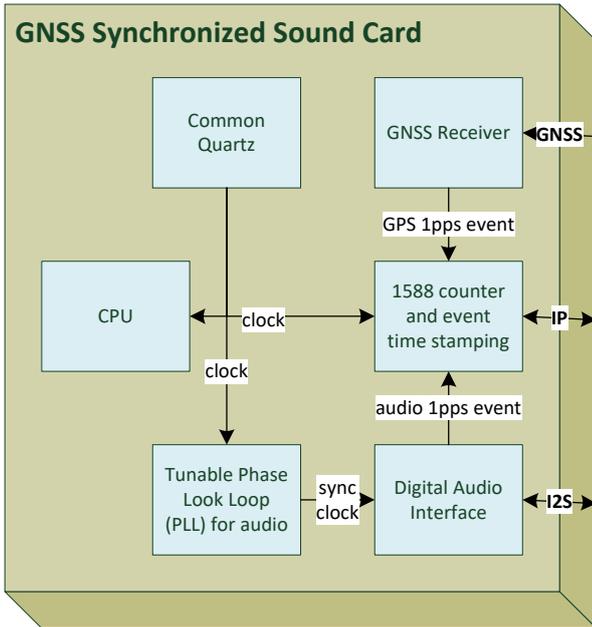


Fig. 1. Block diagram of GNSS synchronized sound card

the described key factors of our theoretical concept. Regarding the upcoming implementation, our concept considers an i.MX7 board by NXP Semiconductors. The i.MX7 series offers a highly integrated processor designed to enable secure and portable applications within the Internet of Things and it suits our demands because it supports the required features with respect to IEEE 1588 and audio word clocking. In that context we will now describe the particular hardware architecture that supports hardware timestamping and time keeping within the Ethernet driver. Afterwards, we will explain how the clocking is realized and how the audio word clock can be controlled and fine-tuned.

#### A. Hardware architecture

To allow for IEEE 1588, the MAC hardware in i.MX7 by NXP Semiconductors is combined with a time-stamping module to support precise time-stamping of incoming and outgoing frames and granule control of the IEEE 1588 time [11]. Figure 1 shows the block diagram of the i.MX7 board architecture.

At the centre of the time stamping module is a 32-bit counter register, which keeps track of IEEE 1588 time on the hardware level. It is incremented after every rising edge received from the oscillator by an amount specified by  $ENET\_ATINC[INC]$  register. In our use case, this value is set to 10 nanoseconds because it corresponds to the Ethernet clock with a frequency of 100 MHz. The  $ENET\_ATPER$  register contains the number of nanoseconds after which the counter will wrap around. It is programmed with a value of  $10^9$  so that the counter resets itself in intervals of one second [11].

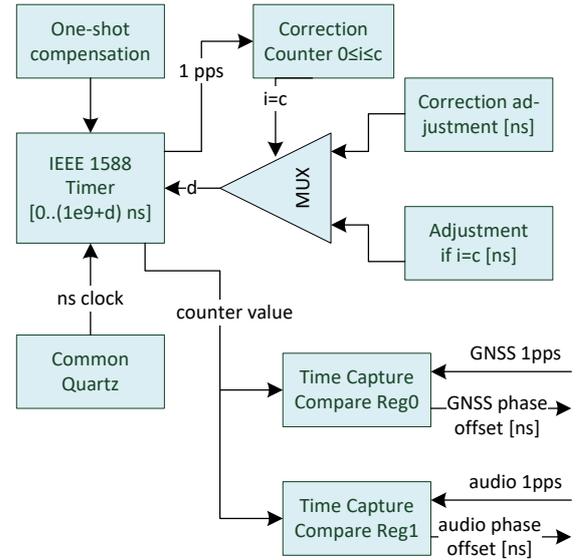


Fig. 2. Block diagram of the adjustable IEEE 1588 counter and the event time stamping

The block diagram of the adjustable timer is illustrated in figure 2.

The  $ENET\_ATCOR$  register is designed for the fine grain tuning of the counter. It defines after how many clock cycles the correction counter should be applied. The amount of correction is specified in  $ENET\_ATINC[INC\_CORR]$ . If the value of  $ENET\_ATINC[INC\_CORR]$  is greater than  $ENET\_ATINC[INC]$  the counter speeds up, if it is less the counter slows down. Furthermore, the system supports a one-shot offset event generation. In that context, the module contains the  $ENET\_ATOFF$  register. It holds the final nanosecond value after which the counter is reset for one single time.

#### B. Audio clocking architecture

The audio word clock in the i.MX7 is derived from the board's main 24 MHz phase-locked loop (PLL) oscillator. In that context, our requirements determine that the audio word clock's frequency must reside in a significantly lower dimension of 44,1 kHz, 48 kHz or 96 kHz and therefore represents a divisor of the board's main frequency of 24 MHz. Furthermore, the final frequency must be adjustable with respect to our previously described synchronization approach. Regarding the fine-tuning functionality the board contains three registers: In the following the  $CCM\_ANALOG\_PLL\_AUDIO[DIV\_SELECT]$  register will be abbreviated with  $Div$ , the  $CCM\_ANALOG\_PLL\_AUDIO\_NUM$  register will be abbreviated with  $Num$  and the  $CCM\_ANALOG\_PLL\_AUDIO\_DENOM$  register will be abbreviated with  $Denom$ . Based on these registers the output of the audio PLL depends on the following calculation:

$$F_{output} = F_{osc} * (Div + Num/Denom)$$

$Div$  multiplies the 24 MHz base clock frequency with the integer specified. The actual granule fine tuning is provided by adding a 32-bit fraction denoted by  $Num$  and  $Denom$ . The resulting audio PLL goes through a pre-divider and a post-divider to get the respective clock frequencies. Pre and post dividers have 64 steps so the frequency change is always an integer multiple of the audio PLL frequency:

$$F_{peripheral} = F_{src} * (Div_{post}/Div_{pre})$$

Afterwards, a so-called Sound Asynchronous Interface root clock (SAI) is derived from the audio PLL, which represents the master for the final audio word clock. The SAI root clock feeds a bit clock generator, which eventually generates the final square-waves-based audio word clock and eventually determines the sample capture and playback of the sound card.

### III. IMPLEMENTATION

This chapter describes the actual implementation of our concept. First, we describe how the synchronization of the local time to UTC time is being realized with an accuracy of one second. Secondly, granule control of the time with nanosecond accuracy is explained and how to keep the clock synchronized with the GNSS time using a PID controller. Eventually, the final audio clock is synchronized with the IEEE 1588 clock, which in turn is in sync with the GNSS clock as intended.

#### A. IEEE 1588 synchronization to UTC time

In order to let the IEEE 1588 clock follow UTC time, first, the local clock must be set to UTC time. This time can be received from the GNSS module periodically once every second [1]. A daemon called *gpsd* is used as an interface between the i.MX7 and the GNSS module. *Gpsd* provides a socket connection between the module and the host [6]. The IEEE 1588 clock ID can be retrieved with the function *phc\_open()* [9]. The definition of this procedure is found in the library *phc2sys* [9]. After the retrieval of the clock ID the POSIX function *clock\_settime (clockid t clockid, const struct timespec \*tp)* [8] is called and the current time is read from the *gpsd* buffer. Hence, the IEEE 1588 time is set to UTC time at the nearest second.

#### B. Granule control of the 1588 clock

Once the 1588 clock time is accurately set at the nearest second, an offset between the start of a second in the UTC time and the local clock time can be observed as described in the concept. The required offset compensation is realized via hardware time stamping. As soon as the 1-PPS pulse occurs on the interrupt line, the current value of the 1588 counter is latched on to the Timer Capture Compare Register *ENET\_TCCR* by the hardware to be inspected later on by the software [11]. This enables a precise calculation of the phase offset. Since the counter is reset in intervals of one second, this value represents the true second offset between the local clock and the absolute start of the second. A one-shot event through *ENET\_ATOFF* is then applied in the

interrupt handler, whose value is set to the value latched in *ENET\_TCCR*, which enables the timer. When the 1588 counter reaches this value, it is set to zero and starts again resulting in the desired offset removal.

After the offset compensation has been applied it is possible to take care of the actual clock drift. The drift can be controlled by an adjustment value that is applied to the counter every second. Ideally, every second would consist of  $10^9$  ns. However, as the clocks are drifting, the GNSS second is not exactly equal to  $10^9$  ns of the i.MX7 quartz. Instead, we see a difference to be equalized. To compensate for this clock drift, the IEEE 1588 timer counts to  $10^9 + d_{normal}$ , where  $d_{normal}$  is the drift offset. This principle compensates the clock drift with a certain degree of precision, however, we further optimize it by using  $d_c$  as a correction value instead of the  $d_{normal}$  value once every  $c$  seconds. This approach is called Proportional-Integral-Derivative (PID).

More precisely, the *ENET\_ATINC[INC\_CORR]* register and the *ENET\_ATCORR* register, which allow granule control over the 1588 counter, can be sped up or slowed down using the Timer Increment. The *ENET\_ATINC[INC\_CORR]* register has the new increment and the *ENET\_ATCORR* register defines its frequency. When the number of clock cycles of the 1588 counter equals the *ENET\_ATCORR* value, the 1588 counter is incremented by the *ENET\_ATINC[INC\_CORR]* nanosecond value instead of the usual *ENET\_ATINC[INC]* value.

The resulting adjustment assumes a difference in time between our 1588 clock start of second and the GNSS start of second. The difference is obtained from the current value latched on to the *ENET\_TCCR* register. Speeding up the 1588 counter is realized by increasing the *ENET\_ATCORR* register value. Decreasing it slows down the timer.

#### C. Synchronizing the audio clock with the IEEE 1588 clock

The final step of synchronizing the audio PLL with the IEEE 1588 clock is realized via a task, which is scheduled in intervals of one second. It receives the clock drift correction from the previous step. According to this value, the audio PLL frequency is changed respectively in order to reflect the change of the audio word clock speed. However, the clock drift compensation of the audio clock is different compared with IEEE 1588: The audio PLL allows to tune the audio bit clock with a very fine grain resolution of less than 0.1 Hz.

Even if the clock drift of the audio signal is compensated, we will need to determine the precise time phase offset of the audio signal with respect to the GNSS clock. Otherwise, the incoming and outgoing audio signals would not be in sync. In order to achieve this, we introduce a special synchronization mode. This synchronization mode is activated only at the startup period because we assume that the time offset does not change if the clocks are running synchronously. Instead of a digital audio output signal, an artificial 1-PPS signal is generated.

In the synchronization mode, the serial audio bit output of the SAI is connected with the serial audio input to a direct

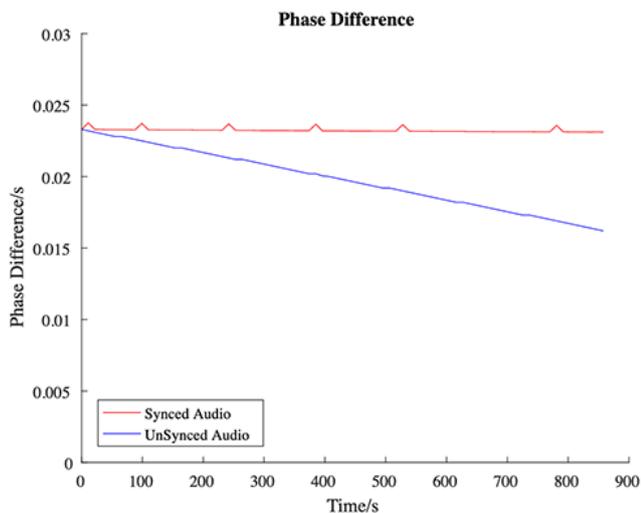


Fig. 3. Audio word clock drift with and without our synchronization

feedback loop. In addition, the SAI data output – filtered by a D flip-flop that is driven by the audio bit clock – is connected to an IEEE 1588 time capture register (similar to the 1-PPS signal of the GNSS receiver). Then, we can measure the delay between sound and GNSS if playing out an artificial digital audio pattern.

#### IV. EVALUATION

The clear and obvious purpose of this paper is the removal of clock drift in remotely distributed sound systems. Therefore we decided to perform our evaluation based on the comparison of two sound card's word clocks with regard to the phase of a low-frequency square wave, which is predestined in that context. In our setup, we generate a 20 Hz square signal with a frequency generator and feed it into two sound cards. The outputs are fed into a 2-channel oscilloscope, which displays both signals at the same time. If the sound card's clocks exhibit a drift the phase of the displayed square waves will drift as well over time. If our implementation is successful we expect the signal to stay in phase and the image on the oscilloscope should stay the same. The duration of the measurement was set to 15 minutes and we retrieved a phase measurement in intervals of 11 seconds. In figure 3 we present our results of this evaluation with and without our developed synchronization approach.

It is clear that without synchronization the phase drift increases proportionally and amounts to 7.5 ms after the final duration of 15 minutes. With our synchronization being applied the phase remains the same because the audio word clocks don't exhibit a drift anymore. In fact we can observe slight phase variations in fixed intervals, however, they are immediately compensated by our applied algorithm and cannot be considered problematic regarding the demands of our described use case.

#### V. CONCLUSIONS AND FUTURE WORK

In this paper, we describe the successful implementation of a GNSS-based sound card synchronization technique for devices distributed in wide-area-networks (WAN) such as the public Internet. Measurement results clearly show the inherent audio word clock drift of approximately 7.5 ms over a duration of 15 minutes without our synchronization technique and precise synchrony of both sound cards and in turn no clock drift when applied. To our knowledge this is the first implementation able to provide the described functionality in the sound card domain. In the near future, we will develop a custom sound card, which benefits from this approach and integrates it into our remote music system. The drawback, however, is that GNSS is not necessarily available in any given environment – especially in basement rehearsal chambers – which is why we will also apply our older approach side-by-side with the new.

#### REFERENCES

- [1] L.J. Arceo-Miquel, Yuriy Shmaliy, and Oscar Ibarra-Manzano. "Optimal Synchronization of Local Clocks by GPS 1 PPS Signals Using Predictive FIR Filters". In: *IEEE Transactions* (2009), pp. 1833–1840. DOI: 10.1109/TIM.2009.2013654.
- [2] Audinate Website. *Dante Overview*. [Online; accessed 12-May-2019]. 2019. URL: <https://www.audinate.com/solutions/dante-overview>.
- [3] Alexander Carôt. "Musical Telepresence – A Comprehensive Analysis Towards New Cognitive and Technical Approaches". PhD thesis. Institute of Telematics – University of Lübeck, Germany, 2009.
- [4] Alexander Carôt and Christian Werner. "External latency-optimized soundcard synchronization for applications in wide-area networks". In: *Proceedings of the 14th regional AES Convention*. Tokyo, Japan, July 2009.
- [5] Bálint Ferencz. *Hardware Assisted IEEE 1588 Clock Synchronization Under Linux*. Master Thesis. Budapest University of Technology and Economics, 2013.
- [6] *GPSd reference manual*. [Online; accessed 12-May-2019]. URL: <http://catb.org/gpsd>.
- [7] Christoph Kuhr and Alexander Carôt. "A Jack Sound Server Backend to Synchronize to An IEEE 1722 AVTP Media Clock Stream". In: *Proceedings of the Linux Audio Conference 2019*. Stanford, USA, 2019.
- [8] Donald A. Lewine. *POSIX programmers guide*. first. O'Reilly, 1994.
- [9] *phc.h Source code*. [Online; accessed 12-May-2019]. URL: <https://github.com/richardcochran/linuxptp/blob/master/phc.h>.
- [10] Ken C. Pohlmann. *Principles of Digital Audio*. fifth. The McGraw-Hill Companies, 2005.
- [11] NXP Semiconductors. *iMx7d Dual Applications Processor Reference Manual*. 2019.

# Palmprint Recognition Based on Convolutional Neural Network-Alexnet

Weiyong Gong  
School of Electronics and  
Information Engineering MOE Key  
Lab for Intelligent Networks and  
Network Security Xi'an Jiaotong  
University No.28 xianning west  
road Xi'an, China  
gong\_w\_y@stu.xjtu.edu.cn

Xinman Zhang  
School of Electronics and  
Information Engineering MOE Key  
Lab for Intelligent Networks and  
Network Security Xi'an Jiaotong  
University No.28 xianning west  
road Xi'an, China  
zhangxinman@mail.xjtu.edu.cn

Bohua Deng  
School of Electronics and  
Information Engineering MOE Key  
Lab for Intelligent Networks and  
Network Security Xi'an Jiaotong  
University No.28 xianning west  
road Xi'an, China  
bohuadeng@qq.com

Xuebin Xu  
Guangdong Xi'an Jiaotong  
University Academy. No. 3,  
Daliangshuxiang East Road  
Foshan, China  
ccp9999@126.com.

**Abstract**—In the classic algorithm, palmprint recognition requires extraction of palmprint features before classification and recognition, which will affect the recognition rate. To solve this problem, this paper uses the convolutional neural network (CNN) structure Alexnet to realize palmprint recognition. First, according to the characteristics of the geometric shape of palmprint, the ROI area of palmprint was cut out. Then the ROI area after processing is taken as input of convolutional neural network. Next the PRelu activation function is used to train the network to select the best learning rate and super parameters. Finally, the palmprint was classified and identified. The method was applied to PolyU Multi-Spectral Palmprint Image Database and PolyU 2D+3D Palmprint Database, and the recognition rate of a single spectrum was up to 99.99%.

## I. INTRODUCTION

WITH the development of network and information technology, the society has put forward higher and higher requirements for the security of information systems. Biometric recognition technology has gradually become one of the important methods to enhance the security and stability of information systems. Biometrics is a technology that uses human physiology or behavioral features for automatic identification [1]. Biometrics are unique personal attributes, which have the characteristics of stability, diversity and individual differences. A number of biometrics have been used, including fingerprint, face, iris, signature,

finger vein, etc. [2]. Meanwhile, palmprint recognition technology is also developing rapidly [3].

Compared with other biological features, palmprint has many unique advantages, and each person's palmprint has different characteristics. Palmprint combines the texture and line features of the palm, and these features do not change over time. In comparison, the collected palmprint is easier to obtain rich personal information due to its larger area than fingerprint. Therefore, palmprint attracts more and more scholars' attention due to its advantages. Traditionally, researchers combine machine learning methods, such as SVM [4], KNN [5], etc., with feature extraction methods, such as LBP [6] and HOG [7], in palmprint recognition. Contrast to machine learning, deep learning achieves automatic feature extraction and classification.

Since 2012, the method based on deep convolutional neural network has achieved remarkable results in various computer vision tasks [9]. In face recognition based on deep learning, researchers have made extensive research [10]. In the aspect of palmprint recognition, researchers have carried out corresponding research experiments with the convolutional neural network method, and achieved good results [11]. In this paper, Alexnet structure is adopted, and experiments and improvements are made on PolyU multispectral Database and PolyU 2D+3D Palmprint Database. The accuracy of each spectrum is quite good, which further proves the effectiveness of the deep learning method in Palmprint recognition.

The rest of the paper is organized as follows: Section II introduces the palmprint preprocessing method and the

---

This work was not supported by any organization

palmpoint recognition method based on deep learning; Section III describes the experimental results obtained by verifying the algorithm on palmpoint database; Section IV concludes the paper.

II. METHODS

A. Image preprocessing

Palmpoint contains rich texture and structural features, such as the main line, wrinkles, triangulation and detail points. When collecting images, images of the same palmpoint collected at different times will have different degrees of rotation and translation, and the size of palmpoint collected at the same time may also be different. Therefore, before feature extraction and recognition of palmpoint, it is necessary to extract the effective ROI area of palmpoint containing main features. The whole processing process is shown in Fig.1. ROI extraction is a key step, and the correct ROI extraction is conducive to image alignment, improving the efficiency of feature matching, and finally giving a positive impact on the recognition results.

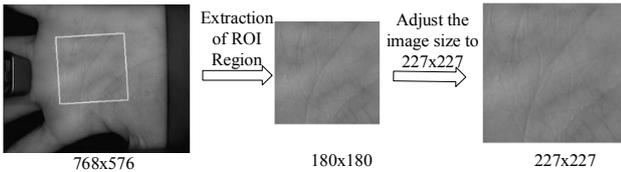


Fig. 1 Image preprocessing

The relative position of the ROI area in the palm is fixed and located in the center of the palm. Zhang et al. proposed edge-based palmpoint positioning processing method [12], which can accurately extract ROI images and thus has been widely used. This article uses this method to process the ROI area with rich features for the following part of the The ROI area of CASIA-Multi-Spectral- PalmpointV1 is shown in Fig. 2

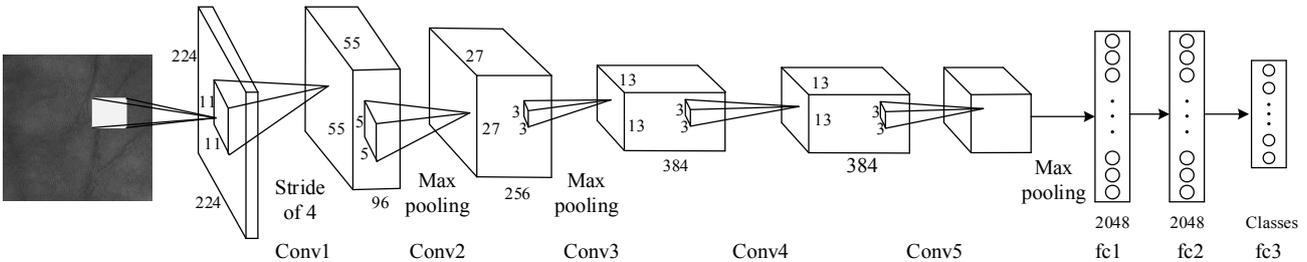
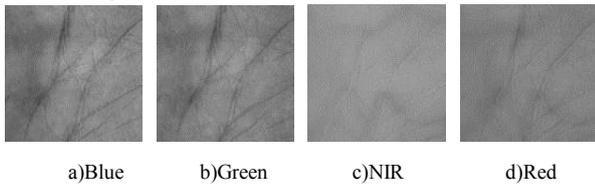


Fig. 3 Structure of Alexnet

Fig. 2 ROI region obtained by image preprocessing

B. Convolutional Neural Network - Alexnet

Convolutional neural network (CNN) is a branch of deep learning which is widely studied and applied. It is not only a multi-layer network model, but also an improvement of BP neural network. They all use forward propagation to output calculated values, and use back propagation to adjust weights and biases. Different from the classical recognition algorithm, CNN repeatedly uses the convolution operation and pooling operation in the original input to obtain increasingly complex feature graphs, and finally directly outputs the results through the full connection. It mainly includes five parts which are input layer, convolution layer, pooling layer, full connection layer and output layer.

In this paper, Alexnet [9] is used and improved. It contains 8 layers of neural network, 5 convolution layers, 3 pooling layers and 3 full connection layers. The structure of the whole system is shown in table 1 and fig.3.

In Alexnet, the author used ReLU activation function and Dropout to improve speed and accuracy. Fig.4 (a) shows the ReLU activation function and the definition is shown in formula (1). Compared with Sigmoid and Tanh activation function, this activation function makes the network converge more rapidly. It can combat the gradient vanishing problem and has high calculation efficiency. When we use ReLU activation function, with the progress of training, there may be a situation where the neurons die and the weight cannot be updated. If that happens, then the gradient through the neuron from this point will always be zero.

$$f(y) = \begin{cases} y, & \text{if } y > 0 \\ 0, & \text{if } y \leq 0 \end{cases} \quad (1)$$

So, He [13] et al. proposed a new nonlinear correction activation function PReLU, whose definition is shown in formula (2). Compared with ReLU, PReLU converges faster. And although PReLU introduces extra parameter "a", it hardly needs to worry about overfitting. In addition, we can achieve learning updates of "a" through back propagation, which enables neurons to select the best gradient in the negative region. So in this article, we use PReLU as an activation function for Alexnet.

$$f(y) = \begin{cases} y_i, & \text{if } y_i > 0 \\ a_i y_i, & \text{if } y_i \leq 0 \end{cases} \quad (2)$$

TABLE I.  
ARCHITECTURE OF ALEXNET

Layers	Details
Input	Image Input(227x227)
Conv1	11x11 Convolutions with stride 4, padding 0
Relu 1	PRelu
Norm1	Cross channel normalization
Pool1	3x3 max pooling with stride 2 and padding 0
Conv2	5x5 Convolutions with stride 1, padding 2
Relu 2	PRelu
Norm2	Cross channel normalization
Pool2	3x3 max pooling with stride 2 and padding 0
Conv3	3x3 Convolutions with stride 1, padding 1
Relu 3	PRelu
Conv4	3x3 Convolutions with stride 1, padding 1
Relu 4	PRelu
Conv5	3x3 Convolutions with stride 1, padding 1
Relu 5	PRelu
Pool3	3x3 max pooling with stride 2 and padding 0
fc1	4096 fully connected layer
Relu 6	PRelu
fc2	4096 fully connected layer
Relu 6	PRelu
fc3	n_classs fully connected layer
softmax	softmax

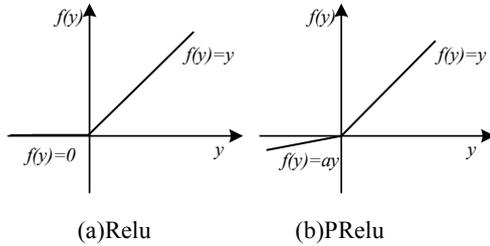


Fig. 4 ReLU vs PReLU

### III. EXPERIMENT

#### A. Public Palmprint Database

Classic Palmprint recognition technology is mainly based on 2D Palmprint, which is convenient to collect and contains rich texture information. Therefore, we use 2D Database in PolyU 2D+3D Palmprint Database. Multispectral Palmprint can obtain more abundant Palmprint features and a higher recognition effect, because use the spectrum of different wavelengths, skin absorption and reflectance is different, you can collect multiple bands of palmprint images, and these images are not easy to be forged. So we used the PolyU Multi - Spectral Palmprint Image Database.

1) *PolyU 2D+3D Palmprint Database*: PolyU 2D+3D Palmprint Database [14] contains 8000 sample images collected from 400 different palms. Samples of each palmprint were collected in two separate sessions, and 10 samples were collected in each experiment. The average time interval between the two sessions was one month. In the experiment, samples taken from the same palm can be regarded as the same class, and samples taken from different palms can be regarded as different classes. Therefore, this sample base contains a total of 400 classes, with each class

containing 20 samples. Each sample contains its 3D ROI (area of interest) and corresponding 2D ROI. This article uses the part of 2D ROI in the database, and the image size is  $128 \times 128$ .

2) *PolyU Multi-Spectral Palmprint Database*: PolyU Multi-Spectral Palmprint Image Database [15] collected 250 people left hand and right hand, a total of 500 classes of sample. Each palm was collected in two separate sessions, each time 6 palmprint images were collected, and the average time interval between the two sessions was 10 days. Each collection was done under lighting conditions of four different spectra, Red, Green, Blue, and Near Infrared, to collect palmprint images of four bands. The specific four bands are near-infrared (band I, wavelength 880nm), red (band R, wavelength 660nm), green (band G, wavelength 525nm), and blue (band B, wavelength 470nm). Therefore, there are a total of  $4 \times 12 \times 500 = 24000$  palmprint images of all palms in each band. Using the above palmprint processing method, the ROI area of  $128 \times 128$  palmprint was obtained for the next experiment.

#### B. Experimental Environment

The training equipment used in this article includes an eight-core Intel i7 processor, an NVIDIA1060 graphics card, and 16GB of memory. The experiment of this paper is carried out in the open source machine learning library Tensorflow2.0 of Google, and the programming language is Python3.5.

#### C. Experiments Results

In order to test and select the appropriate learning rate, we selected 2D Database in the Palmprint Database of PolyU 2D+3D for the experiment. We selected all 400 classes of data in the database. After data preprocessing, we divided the data into 60% training set, 20% test set and 20% verification set to compare the performance.

We tested the influence of different learning rates on the recognition rate with 2500 iterations, and the results are shown in table II. After comprehensive experiment comparison and selection, we found that the convergence rate and training time synthesis reached the best speed and the maximum recognition rate when the recognition rate was 0.0008.

We divided all the data in the database into 60% training set, 20% test set and 20% verification set. The determined learning rate and network structure are applied to conduct training on the PolyU 2d+3d Palmprint Database and PolyU Multi - Spectral Palmprint Database, the results are shown in table III below. In addition, classical methods such as LBP and HOG are used to extract features from the same training set, and classical algorithms such as KNN and SVM are used for training and classification recognition. Then, we use LDA to reduce the dimension of data after LBP and HOG feature extraction, and then conduct classification and recognition. All results are shown in table III below.

TABLE II.  
PERFORMANCE COMPARISON

Learnin-g rate	Iterations	Loss	Time/s	Accuracy of test data	Recog-nition rates
0.01	500	1.5120	448.98	56.00%	99.94 %
	1000	0.0533	1352.85	90.00%	
	1500	0.0001	2461.74	100.00%	
0.001	500	0.7400	436.48	70.00%	99.95 %
	1000	0.0592	1428.42	92.00%	
	1500	0.0001	2561.06	100.00%	
0.0008	500	0.7454	378.42	90.00%	99.96 %
	1000	0.0147	1226.62	100.00%	
	1500	0.0026	2427.54	100.00%	
0.0005	500	0.9277	444.21	46.00%	99.95 %
	1000	0.0719	1242.87	96.00%	
	1500	0.0048	2449.93	100.00%	
0.0003	500	1.5656	441.12	40.00%	99.56 %
	1000	0.1695	1236.46	86.00%	
	1500	0.0025	2431.89	98.00%	
0.0001	500	2.4503	437.07	78.00%	99.64 %
	1000	0.6367	1227.33	74.00%	
	1500	0.3356	2288.35	92.00%	

TABLE III.  
TEST RESOULT

Methods	Recognition rate of Spectras or Databases (%)				
	Red	Blue	Green	NIR	Sub2D
Alexnet	99.99	99.99	99.97	99.86	99.96
LBP+KNN	88.35	89.55	91.44	87.10	90.31
LBP+SVM					
HOG+KNN	97.90	98.40	95.95	96.45	97.31
HOG+SVM					
LBP+KNN	91.60	87.80	88.25	92.75	94.31
LBP+SVM	89.50	87.40	89.10	93.45	93.81
HOG+KNN					
HOG+SVM	99.65	99.83	99.35	99.35	99.84
HOG+SVM	99.60	99.45	98.30	99.10	99.25

It can be seen from the experiment that the method of deep learning can get a better recognition rate in each spectrum and database, which also proves the effectiveness of the method of deep learning in palmprint recognition. At the same time, the experiment also proves that in the aspect of palmprint recognition, the feature extraction and learning ability of deep learning is slightly stronger than the classic algorithm combination of HOG, LBP, KNN and SVM.

IV. CONCLUSION

The classical palmprint recognition method needs a series of tedious operations such as feature extraction, feature selection and classifier selection, and has great limitations in the process of feature extraction and feature selection. Compared with the classical algorithm, the convolutional neural network has the ability to directly input the image and then obtain the classification result, and it has a good

nonlinear fitting ability. When using convolutional neural network for palmprint recognition, we do not need to construct feature extraction algorithm. In this paper, a classic convolutional neural network Alexnet is used for palmprint recognition, and its recognition rate can reach up to 99.99% on test data, which is another idea of palmprint recognition.

ACKNOWLEDGMENT

This work is supported by National Natural Science Foundation (No. 61673316), and Major Science and Technology Project of Guangdong Province (No. 2015B010104002).

REFERENCES

- [1] Jain A. K., Ross A., Prabhakar S, "An introduction to biometric recognition," *IEEE Transactions on Circuits & Systems for Video Technology*, vol.14, pp.4-20, Jan. 2004. doi:10.1109/TCSVT.2003.818349
- [2] Unar J A, Seng W C, Abbasi A, "A review of biometric technology along with trends and prospects," *Pattern Recognition*, vol.47,pp.2673-2688, August. 2014. doi:10.1016/j.patcog.2014.01.016
- [3] Kong A, Zhang D, Kamel M, "A survey of palmprint recognition," *Pattern Recognition*, 2009, vol.42, pp.1408-1418, July. 2009. doi: 10.1016/j.patcog.2009.01.018
- [4] Wu Y P, Tian J W, Xu D, et al. "Palmprint Recognition Based on RB K-means and Hierarchical SVM," *International Conference on Machine Learning & Cybernetics*, 2007. doi:10.1109/ICMLC.2007.4370778
- [5] Kumar A, Bhargava M, Gupta R, et al. "Palmprint Authentication Using Pattern Classification Techniques," *International Conference on Swarm*, 2011. doi:10.1109/CIS.2007.106
- [6] Li, Y.f, and Y. Zhang. "Palmprint recognition based on weighted fusion of DMWT and LBP," *International Congress on Image & Signal Processing*, 2011. doi: 10.1109/CISP.2011.6100392
- [7] Jia W, Gui J, Hu R X, "Palmprint Recognition Using Kernel Spectral Regression Discriminant Analysis and HOG Representation," *International Workshop on Emerging Techniques & Challenges for Hand-based Biometrics*, 2010. doi:10.1109/ETCHB.2010.5559288
- [8] Hong D, Liu W, Jian S, "A novel hierarchical approach for multispectral palmprint recognition," *Neurocomputing*, vol.151, pp.511-521, March 2015 doi: 10.1016/j.neucom.2014.09.013
- [9] Krizhevsky A, Sutskever I, Hinton G E, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, pp.1097-1105, January 2012. doi:10.1145/3065386
- [10] Wang M, Deng W, "Deep face recognition: A survey," *arXiv preprint arXiv:1804.06655*, 2018. doi:10.1109/SIBGRAP.2018.00067
- [11] Jalali A, Mallipeddi R, Lee M, "Deformation Invariant and Contactless Palmprint Recognition Using Convolutional Neural Network," *International Conference on Human-agent Interaction*, 2015. doi:10.1145/2814940.2814977
- [12] Zhang, D.; Kong, W.; You, J.; Wong, M, "Online palmprint identification," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol.25, pp. 1041-1049, Sept. 2003. doi:10.1109/TPAMI.2003.1227981
- [13] He K, Zhang X, Ren S, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," *2015 IEEE International Conference on Computer Vision (ICCV)*. doi:10.1109/ICCV.2015.123
- [14] W. Li, D. Zhang, L. Zhang, G. Lu, and J. Yan, " 3-D Palmprint Recognition with Joint Line and Orientation Features, " *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, vol.41, pp.274-279, March 2011. doi: 10.1109/TSMCC.2010.2055849
- [15] David Zhang, Zhenhua Guo, Guangming Lu, etc., "An Online System of Multi-spectral Palmprint Verification", *IEEE Transactions on Instrumentation and Measurement*, vol. 59, pp. 480-490, Feb. 2010. doi: 10.1109/tim.2009.2028772

# A Contribution to Workplace Ergonomics Evaluation Using Multimedia Tools and Virtual Reality

Roman Leskovský, Erik Kučera, Oto Haffner, Jakub Matišák, Danica Rosinová and Erich Stark  
Faculty of Electrical Engineering and Information Technology  
Slovak University of Technology in Bratislava  
Bratislava, Slovakia  
Email: roman.leskovsky@stuba.sk, erik.kucera@stuba.sk

**Abstract**—The paper demonstrates an application developed to help to evaluate ergonomics of a workplace. Ergonomics of a workplace has enormous impact on employees and their long-term work effectiveness, which causes an interest in this field from employers' point of view. The paper describes and compares several attitudes companies use to set up and evaluate workplace metrics, potential of virtual reality (VR) in the process, VR application proposal, implementation within Unity 3D engine and results achieved with implementation of this proposed solution. Current approaches also include motion tracking for ergonomics evaluation. These technologies are often far over smaller companies' budget. Described solution is reasonably priced also for small companies, using cheaper motion capture equipment.

## I. INTRODUCTION

AT THE present time, the interest in health of employees performing monotonous and repetitive routines in industrial companies is increasingly being pursued. One of the new issues is the evaluation of workplace ergonomics in such enterprises. This topic is strongly supported, especially in Germany, where trade unions have an important influence. The paper deals with the proposal of methodology for ergonomics evaluation using modern information-communication technologies (ICT), such as virtual or mixed reality. This vision is fully in line with the European industry revolution - Industry 4.0 [1].

The aim of our work was to contribute to the development of a comprehensive system (application) for the evaluation of workplace ergonomics for medium and small enterprises. As the development of such a system (application) is a complex task that requires a multidisciplinary approach, the work has set out subtasks that relate to computer support for ergonomics evaluation. The project uses a new motion capture suit named Perception Neuron. The advantage of this solution is a system that is less expensive than current solutions. Another benefit is the use of a virtual reality in which individual workplaces can be composed and, if necessary, used for worker training [2].

## II. ERGONOMICS

Ergonomics is the process of designing and deployment of workplaces - objects and systems in an environment so that their layout fits the people who use them. It can be applied to

all processes and locations that include people - workplaces, sports, leisure activities, but also safety and health.

The goal of ergonomics is to refine the layout of environment to minimize the risk of injury or other types of health problems.

Enterprises are concerned with ergonomics because they realize that the better the environment is adapted to the needs of the person, the higher the productivity of person's work is. At the same time, companies try to minimize the costs of illnesses and wounds resulting associated with work at the workplace.

### A. Ergonomics Evaluation

Several disciplines are used to evaluate ergonomics [3], [4]. In the factories, especially in our site, the implementation is predominantly tabulated, with the company prescribing workplace deployment standards, which describe the minimum / maximum table height, desktop size, distance between objects, and so on.

Some of the companies also use innovative technological solutions that are available today. Motion capture methods have been used at the sport sector. Similarly, there are car manufacturers that evaluate the cockpit ergonomics with the aforementioned methods, which record, for example, the movements of people during boarding and disembarking, which are later evaluated.

We attempt to contribute in this area by incorporating current technologies, using motion capture methods in combination with virtual reality [5].

### B. Existing Solutions

The paper [6] lists the basic ergonomic aspects of workplace design. The advantages of the ergonomic design of the working environment are proven by increasing work performance and reducing errors. A turnover of staff is also reduced. Workers are more satisfied at work and have no reason to look for a new job. The benefits can therefore be seen for both organizing the production and the employee:

- 1) For the production organization

- a) reduction of incapacity for work and occupational diseases
  - b) performance improvement
  - c) reducing error and confusion
  - d) improving the mental status of the worker
- 2) For the worker:
- a) improving the mental and physical condition of the worker
  - b) minimizing the signs of mental and physical fatigue
  - c) social benefits - improved self-realization

As stated in [7], the use of virtual reality technology in solutions aimed at increasing workplace ergonomics is an unavoidable trend. Virtual reality applications have these benefits:

- user can enter and walk through the scene on different tracks
- events happen in real time
- scene and objects have 3D character
- scene is not static and objects can be manipulated

Virtual reality models allow:

- replacing physical prototypes with virtual ones
- simulation of the different stages of development in a virtual environment
- improvement and acceleration of product development processes
- scene is not static and objects can be manipulated

The main advantages of human simulation in the 3D environment are:

- shortening development time
- reducing development costs
- improvement of quality and safety
- increase of competitiveness

The use of virtual reality in the design of workplace ergonomics will provide the following functionalities:

- a clearer design of the new work cell
- evaluation of existing assembly line, increasing its efficiency considering the human factor
- energy expenditure for the operation

Tecnomatix Jack is frequently used for ergonomic workplace design [8]. Though, it does not provide a built-in solution that uses VR and its unquestionable benefits. Existing solutions for employee ergonomics analysis in VR are mostly developed by big companies. These software and hardware solutions are often developed for automakers [9].

So far, there is no affordable solution for small and medium-sized companies. For example, these are commercially available virtual reality headsets and affordable motion capture suits. Our proposed solution described in this paper is priced at a maximum of thousands of USD / USD. Existing solutions for large companies are priced at hundreds of thousands.

### III. PROPOSED VISION

The impulse for the project was the idea of connecting motion capture methodology with the visualization capabilities

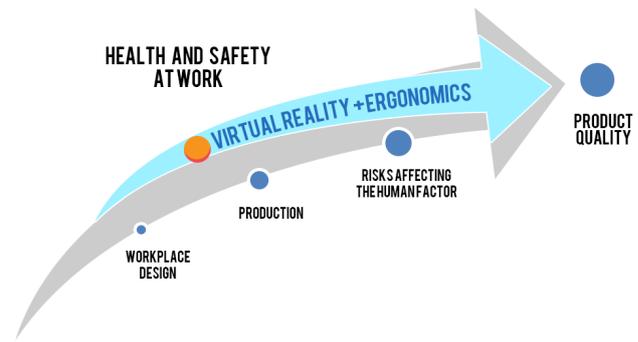


Fig. 1. Impact of involving innovative procedures to enhance product reliability [7]

of VR applications [10] [11], giving users new perspective to the spatial body movement in a virtual workplace.

Recording worker's movements via motion capture technology and then providing users with tools for controlling recorded live movement data similarly as a movie player with functions to control play speed or pause whole moment and also providing tools for modifying environments in way of building custom workplaces, would granted users with full control of those aspects which would make it easy to recognize critical spots of the workplace.

Application use is intended mainly for those kinds of workplaces, where inappropriate ergonomics may cause an increase of wounds. There are connected with manual repetitive actions, that are common e.g. in factories, so it is compulsory that it includes industrial equipment, that could be used to build industrial workplaces. Considering the fact that application should be reasonably priced also for small companies, proper support of inexpensive variants of technologies is required.

### IV. MAIN ASPECTS OF PROPOSED APPLICATION

Based on the vision, our application was proposed. The proposal was divided into several different independent parts, each dealing with different problematics. Application was divided into:

- Supported technologies
- User input / movement
- Environment
- User interface
- Workplace building
- Work simulation
- Limitations

#### A. Supported Technologies

As it was stated in the proposed vision, the application should offer virtual reality environment, which includes the use of a VR headset. It relies on use of a motion capture technology and 3D engine which is a powerful environment when dealing with 3D rendering [12].

There are a lot of VR headsets to choose from. Support for more than one device often meant device dependent development and writing code for each device separately. It was caused by differences in controllers and input. Some of the headsets provides only gestures, others one or two controllers of different types.

Acer WMR (Windows Mixed Reality) was the best candidate for the application and for testing of the proposed system. Its system requirements and performance tests were better on our computers than it was with the other virtual reality headsets considered for use (including HTC Vive, Oculus Rift). Advantages of this headset is lack of external sensors for tracking of headset’s orientation. Tracking sensors are internal – in a form of two cameras detecting movement of the device relative to the room or other surroundings by the changes of positions of referral points in the environment in each image. In addition to performance and saving one HDMI port and not using external sensors, also the price is lower than the price of the other devices. Headset belongs into the family of Windows Mixed Reality headsets. These devices share uniform controllers, which makes the application compatible with each type of headset of this type.

For motion capture technology, price was the key feature that matters as this could be the most expensive part of the whole project. Also used tracking technology was important as motion should be captured right in and during real work process. Some of the workplaces could provide only very constrained setting, making it impossible to capture movement with cameras detecting markers on worker.

Considering mentioned facts, Perception Neuron was selected. Developed by Noitom, this motion capture suit consists of individual sensors called Neurons. Each neuron houses an Inertial Measurement Unit with a accelerometer, gyroscope and magnetometer [13]. This units should provide decent measurement values with reasonable price. Suit allows full body tracking. Noitom also provides a SW in the form of environment, where editing, real-time preview of animations and export to various formats are possible. The only limitations are warnings about wearing suit near the electric devices, which is not always possible in a factory.

*B. User Input and Movement*

A design of user input was not an easy task. It is always important to realize, what are the actions that we need to cover and how many different types and keys will be available for the users.

Using WMR controllers, we got four buttons (one reserved for return to Windows), one joystick and one touchpad per controller. Considering left and right hand-oriented users, it may cause difficulties, if each of the buttons on the left and right controller had different functions. It would be also not possible to completely control the application with just one controller. Because of this, functions of buttons on both controllers will be mirrored. This left only five input keys for genuine actions as buttons on the left controller should call the same actions as the opposite buttons on the right controller.

As defined in the vision, main actions could consist of:

- movement
- displaying/hiding the menu
- selecting items in menu / environment
- confirming action
- playing the simulation
- changing simulation play speed

Reserving joystick for movement works great and feels natural as axis of joystick movement corresponds to the direction of movement in a virtual environment. Touchpad was reserved for scrolling in menus on vertical axis and changing play speed on horizontal axis.

As a specific method of location input, application will use controllers as pointers used to select options in menu panels or position in environment.

All buttons setup proposal is described in Fig. 2.

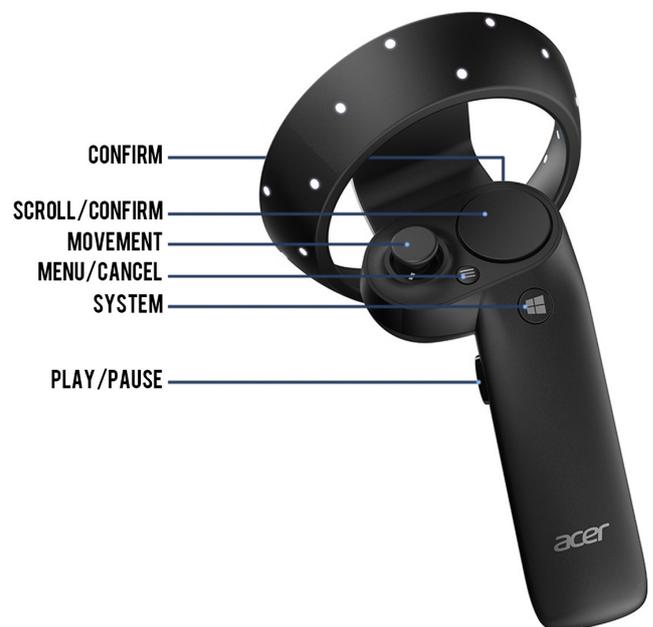


Fig. 2. Controller key bindings

Movement itself was a problem since of appearance of first virtual reality applications. Movement caused disorientation and nausea. Now we know, that this was caused by unnatural movements in virtual environment that human brain had problem to process. Most of the problems were caused by absention acceleration and deceleration forces present while moving in real life. This would be impossible to do by code, as every person’s movement pattern is different from others. Therefore, modern VR applications relies on real physical movement that is detected through inside or outside sensors and then it is transferred into the VR, which is more natural and not causing problems like older applications.

Natural movement is often limited by real surroundings, as user may need to move in larger area but that’s not possible in real setting. In this case application provides procedure of teleportation, where user can change location instantly for

longer distances by pointing at the new location and pushing joystick forward. When user releases the joystick, application moves him/her to the new location.

Application also provides "step back" function, used go back and turning joystick left or right rotates user in desired direction.

### C. User Interface

User interface (UI) should consist of selection menus or panels with additional information to user's visual experience.

Since UI will be presented in VR, there are specific rules for text to be easily readable and provide understandable options.

Important text information should be kept in user's field of view (FOV) as it is something users should be aware of in whatever direction they are looking at. Displayed panels should be far enough, not blocking the view, but fitting in. The size of font of every text information should be large enough so users will be able to read it. Aiming at the buttons should not be challenging task to feel selection more natural, achieved by selecting proper size of buttons.

Providing information in more different panels step by step also works better for VR than displaying everything all at once.

Curved menus (Fig. 3) for bigger panels are preferred unlike straight menu panels where content on the sides is in greater distance, which results in unequal feel of presented options from side of user.

Curve angle should not be too high to avoid claustrophobic feeling. This could be done by moving the center of an imaginary circle, where lies curved menu panel, behind user's back.

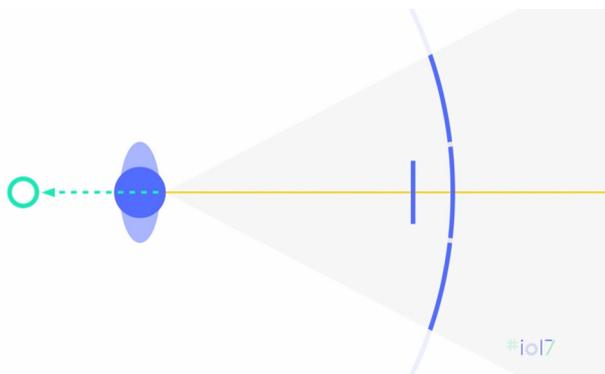


Fig. 3. Layout of curved menu with circle center shifted behind users back [14]

### D. Environment

Environment needs to provide enough space for users to simulate physical body movement actions and also free space where custom workplaces will be built.

Both indoor and outdoor scenes (Fig. IV-D) can be used for the simulation. Indoor scenes would need correct lightning settings and outdoor scene raised skyline to make scene look infinite. Illusion of no boundaries improves user experience (UX).



Fig. 4. Outdoor / Indoor scene examples

## V. WORKPLACE BUILDING

Building of a scene in the application will be a process, where user selects an object from a set of provided 3D models and places these objects to desired location in order to build either a replica of existing workplace or to create and test a new workplace.

Provided items will be sorted into categories to make selection faster. Each item should provide detailed look before selection. Item position in environment will be selected by pointing a controller and pushing button for accept (Fig. 5).



Fig. 5. Proposal of item selection menu

## VI. WORK SIMULATION

Simulation consists of playing recorded body movements of a worker during his/her work.

Users will be able to load film box format (.FBX), which, in addition to polygon and material data of 3D model, also contains animation data. This data will be automatically processed and mapped to 3D engine's animation component.

Loaded animation will be bound to the 3D model of a human figure used as a mannequin. Playing the animation will move mannequin accordingly.

The applications will also provide a control mechanism for control play speed of an animation based on music or video players (Fig. 6).

### A. Limitations

Setting up limits is important to maintain performance of the application.

This could be done by restricting of number of objects that can be placed in single workplace. The application should also focus on polygon optimization of 3D models. Reducing

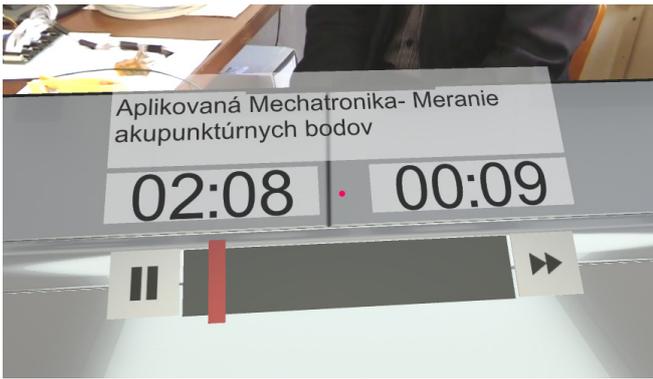


Fig. 6. Video player as an idea for animation player

polygon count in scene will reduce calculation count needed for rendering. Polygon optimization is useful when working with 3D models from CAD. CAD models provides too many details that are not needed in virtual reality visualization (Fig. 7).

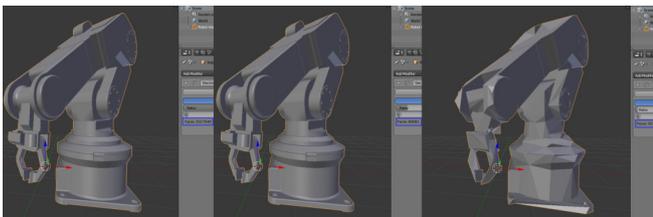


Fig. 7. Example of reducing number of polygons of a 3D model

Another restriction may be setting up a movement area boundary, which defines location, where user can freely move and those places, that are not accessible.

VII. IMPLEMENTATION AND RESULTS

The vision of application was implemented and built by Unity 3D engine. Unity provides build for a lot of platforms including Windows, Mac, Linux, Android, iOS, etc. WMR headsets are bound to operating system Windows as it is the only operating system supported by these devices.

Unity 3D engine is perfect choice for this type of application. The 3D engine has tools and components created for use with virtual reality. Great advantages of this 3D engine are low learning curve, easy scripting system and simplicity of engine’s GUI.

A. Scenes

The application itself is divided into two scenes. The first scene provides simple menu with options of loading animations or workplaces while the second scene is used for building and simulation.

Menu-selection scene is a starting point, where user can load animations and open workplaces. This is a small scene where users find themselves inside a room with reflections to optically enlarge surrounding area.

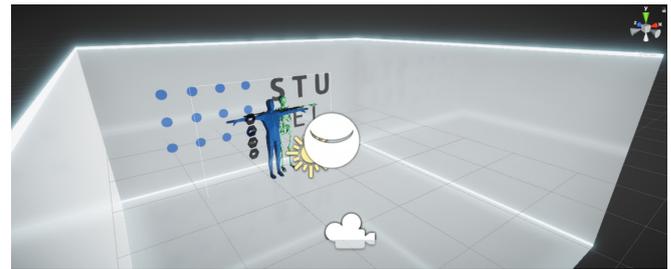


Fig. 8. Main menu scene

Loading of animations is done via system file browser panel, which has similar functions to the traditional Windows file browser. Selection is confirmed by aiming at desired option and clicking the confirm button on virtual reality controllers.

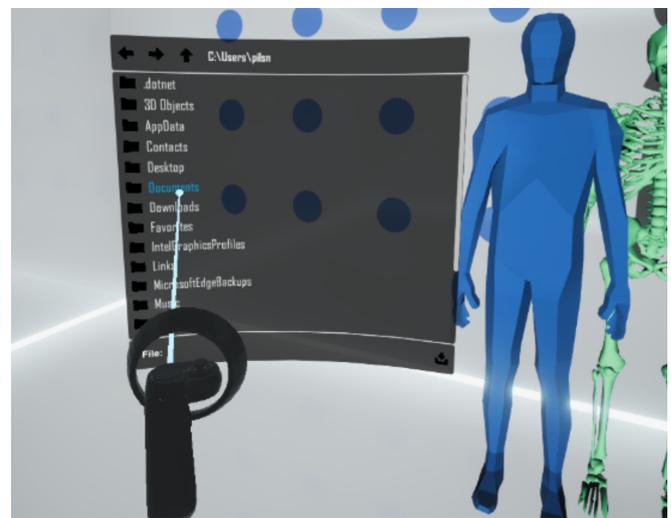


Fig. 9. File browser

Second scene is loaded after animations and workplace are selected. User spawns inside a hangar building. This scene serves for building virtual replicas of actual workplaces. Hangar should provide enough space also for larger workplaces. The door on one of the sides of the building are opened.



Fig. 10. Simulation scene

Outside area contains grass and trees. Also bird sounds can be heard. Even though outside area is not accessible, it creates a relaxed atmosphere.



Fig. 11. Outdoor environment of the simulation scene

### B. Controllers

During application runtime, virtual models of controllers are being rendered for users, if they are turned on and paired with computer. Virtual controllers increase amount of interactivity. Selection ray is displayed with a dot on its end which works as a mouse cursor. Aiming and hovering over buttons and objects with scripted actions will result in selecting these components.

#### REAL CONTROLLER



#### 3D MODEL



Fig. 12. Real and virtual controller

Movement realization is done exactly as proposed. The application provides four types of movements:

- Real movement projected into application – whatever move user do with headset same change is performed by in-game camera
- Teleport – pushing joystick forward and aiming at the ground, users can teleport to a new location
- Step back – pushing joystick back will result at camera moving back a little
- Turn left/right – turn by 30 degrees by pushing joystick left or right

While requested teleporting by pushing joystick forward, markers to indicate ongoing action are displayed to the user.

### C. Raycasting

Method of raycasting is used for collision detection. During detection a ray is cast in a certain direction for specified length and first found interception is returned as a collision point.

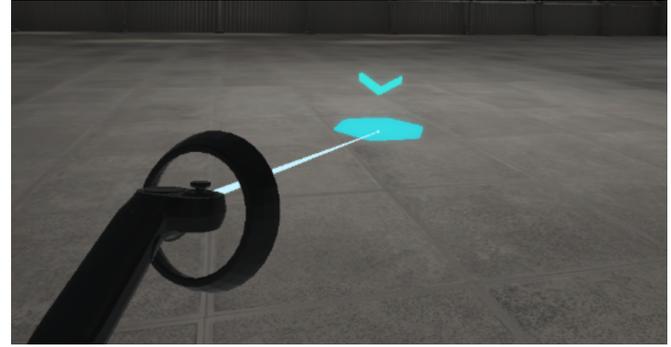


Fig. 13. Teleportation sign

In this case, rays are cast from controllers, checking for editable objects and menu item detections. If confirm button is pressed and user is aiming at such object or menu item at that moment, its action is called in form of C# method written in abstract editable object class overridden for specific use of specific object user is currently clicking on.

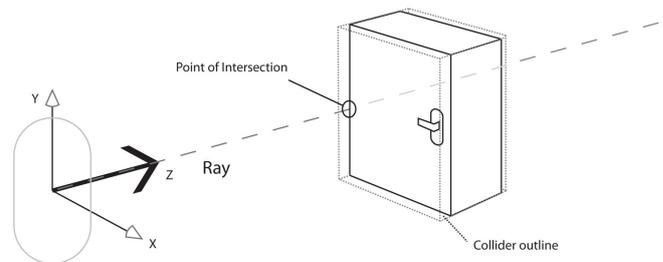


Fig. 14. Raycasting method [15]

As some of the meshes can have too many polygons and complex shapes, colliders are used to simplify these shapes and to receive the collisions. This approach reduces the quantity of calculations needed, although collisions are not always that precise as simplified shape cannot always perfectly align complicated shape.

In Fig. 15, green lines represent edges of a chair collider around its mesh. Colliders at the top of the objects are defining mesh better than those at the base. Reason is that higher parts of collider are more likely to stand in front of other object from user's perspective, which causes blocking rays and detecting object at the front when aiming its direction. When these colliders are better defining its mesh, change-over the objects while moving the controllers is more accurate, representing what user really see. At the bottom areas, while aiming at the ground, there is unlikely that collider could block another object, so those can be simplified so that they don't correspond with actual mesh.

### D. Object Selection And Building

Building of a workplace is done by selecting an object by object from application menu and placing and positioning the object to fit user needs. Objects provided are categorized.

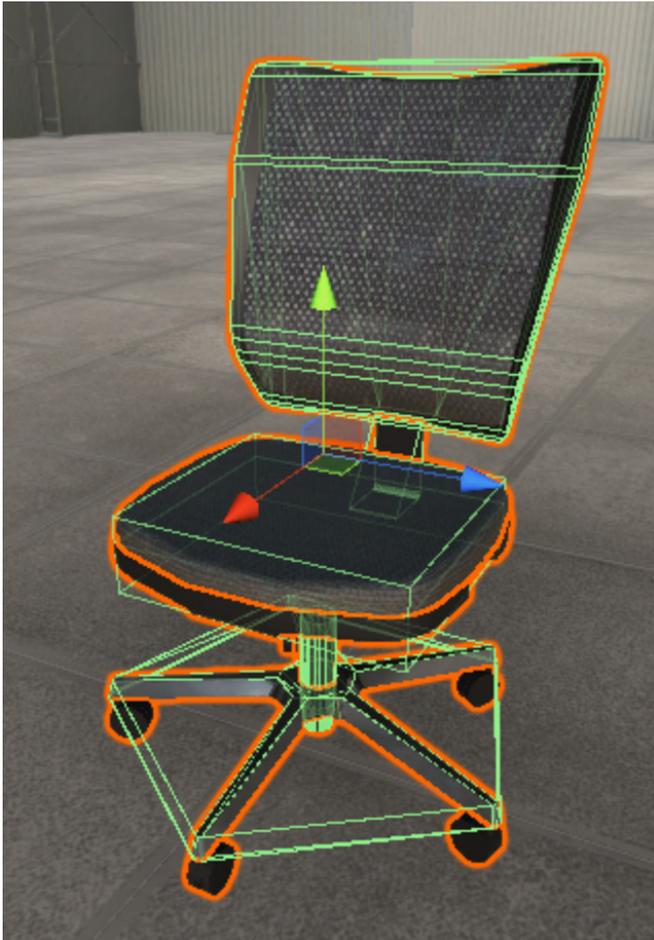


Fig. 15. Object collider example

Menu provides these categories as buttons. Each category contains several optimized objects prepared for use. Items are loaded automatically from application resources based on the folder names. After clicking on category button, objects of selected category are displayed to the user.

Provided categories are:

- Machinery
- Conveyor belts
- Racks
- Tables
- Chairs
- Cabinets
- Electronics
- Other – other objects that do not belong to any of mentioned categories like decoration objects, boxes, walls, etc.
- Primitives – if users cannot find suitable object of those provided by the application, they can replace it with this simple primitive 3D objects as cube or sphere
- Functional – this section contains physical animation player that can be used to control the movement simulation. Category was also prepared for objects like doors

- or elevators that was not implemented yet
- Mannequin – contains humanoid 3D model

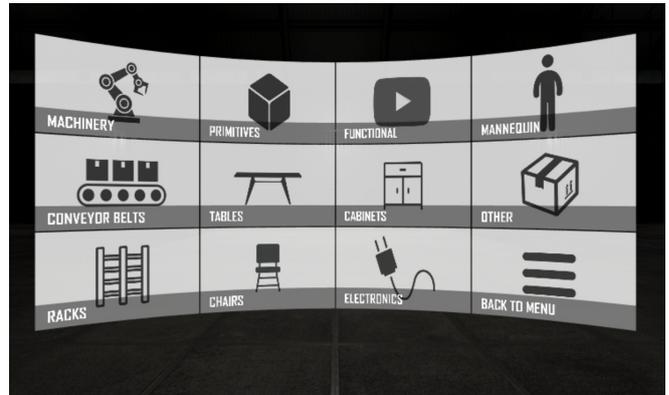


Fig. 16. Build menu

Providing limited number of objects users can use to build their workplaces, it can be hard to create something that looks similar to the real environments. If application contained only very specific 3D models, it would be impossible task.

Due to this cause, whole specific assembly lines were split into smaller reusable parts that can be used multiple times with different setup (size or colour). Doing this, users create their own specific solution themselves using the principle of modularity (Fig. 17) by combining several objects in many ways.

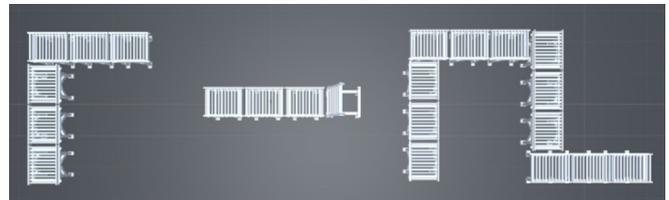


Fig. 17. Modularity principle

### E. Object Detail

When object category in build menu is selected, objects that belong to the category are displayed to the user. There are several objects in each section which could get confusing when showing all at once, especially when also visual of every object has to be displayed. Instead, application presents one object at a time, providing name of the object and detailed look. Arrows on both sides of the object detail can be used to cycle through all objects.

Preview of each object is rendered dynamically. The detail canvas contains a render texture, which can be used to display an image from a camera. Whenever user opens a selection menu, selected item is placed under the whole scene in front of prepared detail camera, where user cannot see it. This camera renders an image, that is projected in the detail canvas of selection menu. When changing the object, old object selected

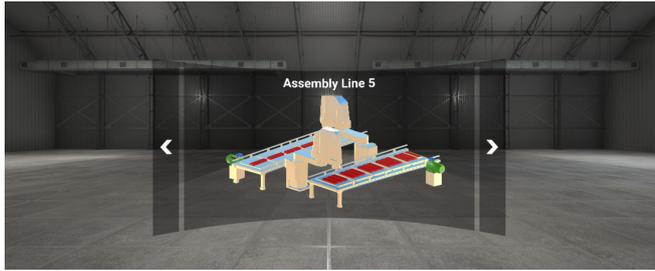


Fig. 18. Selection menu

before is deleted and replaced by the new one just being selected. This will result in change of camera view, which also renders whole new image on the menu canvas.

Detail camera uses orthogonal view, which makes 3D object look more like an image.

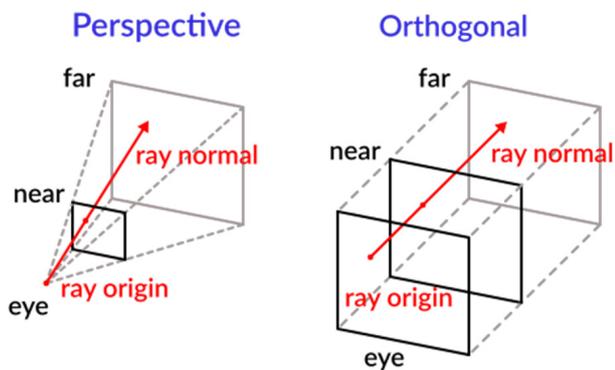


Fig. 19. Orthogonal camera render principle [16]

User confirms the selection by clicking the object image in the canvas, which results in closing the menu and sticking the object to the end of the ray casted from controller, changing its position whenever controller is moved and casted ray collides with the ground. At this moment, object can be rotated moving the joystick right or left. Aiming at desired position and by clicking the submit button, object stops following the ray and becomes solid part of the virtual environment.

#### F. Object Set-Up Options

After selecting the object and confirming its position it stays still. But sometimes it is necessary to be able to change object's rotation or position again or delete the object after its building.

All possible modifications of each object are presented by the setup menu. This menu can be accessed by aiming at the built object and pressing the submit button. When aiming at the object, name of the object and the action are automatically displayed to let user know about the action.

Setup menu provides numerous object settings:

- Change rotation – buttons can be used to change the rotation of placed object. Rotation is changed on three euler axes.

- Change scale – rescaling the object on three axes.
- Materials – preview of all mesh materials – clicking the material will open a material menu
- Detail – similar to the selection menu, creates a copy of modified object. All of the modifications are done affecting only this copy and only by clicking the confirm button copy replaces old object in scene. If not confirmed and cancel/back button is pressed, menu closed and object copy is deleted, preserving original object.
- Buttons
  - Confirm – confirms the changes
  - Move – closes the menu and sticks the object to the end of the ray, user can move object again to its new location
  - Delete – deletes object from scene

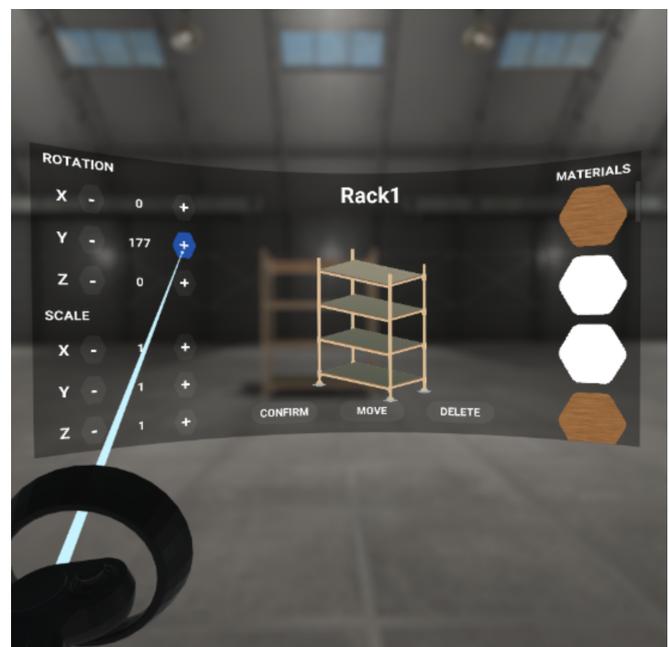


Fig. 20. Object editing menu

Selecting the material from the setup menu opens other menu with all the materials grouped by their type. Clicking on new material will replace the original one. All the materials can be changed. Provided material groups are:

- Simple color
- Fabric
- Glass
- Metal
- Plastic
- Stone
- Wood

Setup menu provides only few options to customize objects, however it can be used to produce totally different object with the identical base 3D model. As an example, here is a cabinet that was rescaled, and materials were changed.

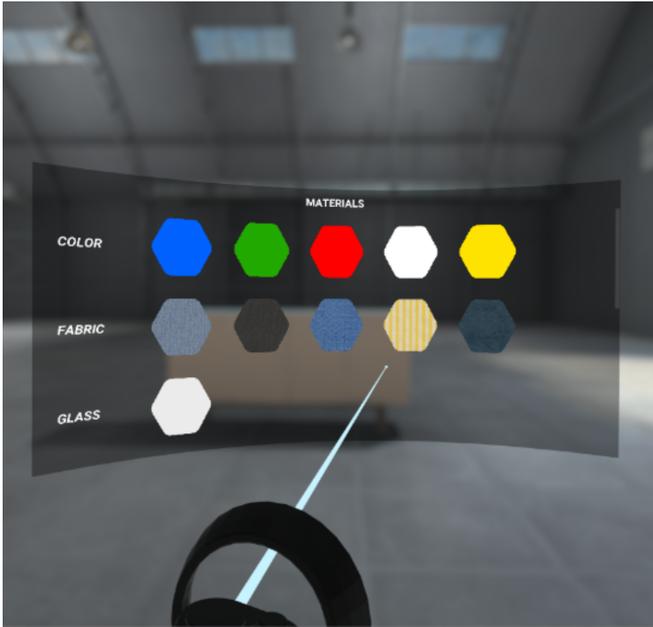


Fig. 21. Material selection menu



Fig. 22. Example of original and edited object

*G. Motion Simulation*

The application provides tools for the simulation which is in this case done by playing recorded motion data as an animation bound to the humanoid 3D model. At first, motion data needs to be captured using motion tracking technology. Data have to be stored as .FBX format which can describe also animation data stored in time frames. This file is then loaded through the application using the file browser.

Unity 3D engine does not provide tools for runtime load of an animation or model to the application (except for asset bundles, which is not suitable for this solution, due to the need of the Unity editor to create a bundle), so an external library is used and included to the application.

Asset Import library is an open-source library dedicated for the import of multiple 3D model data formats into the form, it can be easily processed in third-party applications.

With use of mentioned library, animation data is extracted from loaded .FBX file, stored into animation time frames that form whole animation. Each frame has information about positions, rotations and scales of animated object at a specific moment in time. Changing object position and rotation according to this information continuously, frame by frame,

will create a visual animation. After animation was loaded, user is informed about animation statistics like total number of frames, framerate (frames per second) and duration of loaded animation. Then users may select an environment in which simulation takes place.

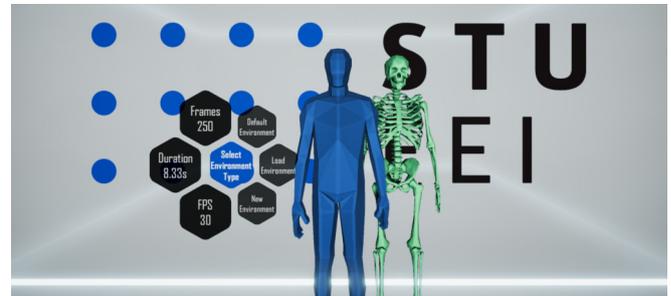


Fig. 23. Animation information after load

In environment, animation is automatically bound to the mannequin – 3D humanoid. To perform the simulation, its necessary to play the animation, which can be done by the controllers as there are keys for play/pause as well as controls for controlling the play speed. In section “functional” of build menu, users can also find physical controller composed of buttons. This controller provides additional information about play speed, animation name, duration and actual time and buttons with the same functionality as described for VR controllers.



Fig. 24. Motion simulation

*H. Results*

Many kinds of environments can be built within the proposed application using / editing provided objects.

VIII. CONCLUSION

The application was developed to help evaluate ergonomics of workplaces with focus on manual repetitive monotonous working activities. One of the main goals was to make the application affordable for small and medium enterprises. The application allows use of affordable motion tracking solutions, which ensures a price drop.



Fig. 25. Office workplace example



Fig. 26. Industrial workplace example

Main purpose of this application is to provide an alternative original look at the working process, create new points of view using the virtual reality technology and provide tools for managing environment and working process in it.

The application can also be used for workplace adjustments and modifications to fit the ergonomics standards or to change design of evaluated environment to solve its deficiencies. It can also be used for designing entire new environments and in some limited way for training of new employees.

The application is not evaluating ergonomics itself, as it needs specialist in this field to do the evaluation. Automatic machine evaluation of workplaces is one of the improvements we are considering to the future. There is also software like Biomechanics of Bodies, that can automatically evaluate ergonomics of human body based on human joints positions and rotations, counting muscle stress and much more. Using software like this, the application could provide users with evaluated data, directly displaying critical human body parts or exact muscles that suffer the most and marking critical area where unnatural movements occur.

Other improvements can provide more user actions and customization to make it suitable for employee training and augmented reality support could bring evaluation to the real workplace.

## ACKNOWLEDGMENT

This work has been supported by the Cultural and Educational Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic, KEGA 030STU-4/2017 and KEGA 038STU-4/2018, by the Scientific Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic under the grant VEGA 1/0733/16, and by the Young researchers support program, project No. 1328 – KVPRI (Quality Control of Production Processes with Augmented Reality in Industry 4.0) and project No. 1327 - VTOVI (Virtual Training of Production Operators for Industry 4.0).

## REFERENCES

- [1] T. Lojka, P. Satala, J. Mocnej, and I. Zolotová, "Web technologies in industry hmi," 09 2015. doi: 10.1109/INES.2015.7329647 pp. 103–106.
- [2] J. Filanova, "Application of didactic principles in the use of videoconferencing in e-learning (in slovak)," in *Innovation process in e-learning*. EKONOM, March 2013. ISBN 978-80-225-3610-3 pp. 1–7.
- [3] D. Consulting. (2014) What is ergonomics? [Online]. Available: <https://www.ergonomics.com.au/what-is-ergonomics/>
- [4] J. Krišťák. (2017) Ergonomic workplace layout (in slovak). [Online]. Available: <https://www.ipaslovakia.sk/sk/ipa-slovník/ergonomicke-ustporiadanie-pracoviska>
- [5] S. Hashimura, H. Shimakawa, and Y. Kajiwara, "Automatic assessment of student understanding level using virtual reality," in *Proceedings of the 2018 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems, M. Ganzha, L. Maciaszek, and M. Paprzycki, Eds., vol. 15. IEEE, 2018. doi: 10.15439/2018F268 pp. 39–45. [Online]. Available: <http://dx.doi.org/10.15439/2018F268>
- [6] A. Lešková, "Ergonomic aspects of workplace design (in slovak)," *Transfer inovácií*, vol. 7, 2004.
- [7] M. Hovanec, "Progressive methods in workplace optimization (in slovak)," *Transfer inovácií*, vol. 25, 2013.
- [8] J. Šesták and J. Čuchranová, "Design of ergonomic construction manufacturing systems for manual assembly (in slovak)," *Transfer inovácií*, vol. 25, 2013.
- [9] K. Procházková, "Ergonomics in vehicles of Škoda auto (in czech)," 2014.
- [10] K. Zhang, J. Suo, J. Chen, X. Liu, and L. Gao, "Design and implementation of fire safety education system on campus based on virtual reality technology," in *2017 Federated Conference on Computer Science and Information Systems (FedCSIS)*. IEEE, 2017. doi: 10.15439/2017F376
- [11] J. Majerník, M. Madar, and J. Mojziso, "Integration of virtual patients in education of veterinary medicine," in *2017 Federated Conference on Computer Science and Information Systems (FedCSIS)*. IEEE, 2017. doi: 10.15439/2017F134
- [12] S. Paszkiel, "Control based on brain-computer interface technology for video-gaming with virtual reality techniques," *Journal of Automation, Mobile Robotics and Intelligent Systems*, vol. 10, no. 04, pp. 3–7, 2016. doi: 10.14313/JAMRIS\_4-2016/26
- [13] H. Kim, N. Hong, M. Kim, S. Yoon, H. Yu, H.-J. Kong, S.-j. Kim, Y. J. Chai, H. Choi, J. Choi, K. E. Lee, S. Kim, and H. Kim, "Application of a perception neuron® system in simulation-based surgical training," *Journal of Clinical Medicine*, vol. 8, p. 124, 01 2019. doi: 10.3390/jcm8010124
- [14] G. Developers. (2017) Designing screen interfaces for vr (google i/o '17). [Online]. Available: <https://www.youtube.com/watch?v=ES9jArHRFHQ>
- [15] T. Lintrami, *Unity 2017 Game Development Essentials*, 3rd ed. Packt Publishing, Jan. 2018.
- [16] J. Linietsky and A. Manzur. (2018) 3d ray casting from screen. [Online]. Available: <http://docs.godotengine.org/en/3.0/tutorials/physics/ray-casting.html>

# Information theoretical secure key sharing protocol for noiseless public constant parameter channels without cryptographic assumptions

Valery Korzhik, Vladimir Starostin,  
Muaed Kabardov, Aleksandr Gerasimovich,  
Victor Yakovlev, Aleksey Zhuvikin  
The Bonch-Bruевич Saint-Petersburg State  
University of Telecommunications,  
Saint-Petersburg, Russia.

Email: val-korzhik@yandex.ru, : star\_vs\_47@mail.ru

Guillermo Morales-Luna  
Computer Science Department  
CINVESTAV-IPIV,  
Mexico City, Mexico.  
gmorales@cs.cinvestav.mx

**Abstract**— We propose a new key sharing protocol executed through any constant parameter noiseless public channel (as Internet itself) without any cryptographic assumptions and protocol restrictions on SNR in the eavesdropper channels. This protocol is based on extraction by legitimate users of eigenvalues from randomly generated matrices. A similar protocol was proposed recently by G. Qin and Z. Ding. But we prove that, in fact, this protocol is insecure and we modify it to be both reliable and secure using artificial noise and privacy amplification procedure. Results of simulation prove these statements.

**Index terms:** key sharing protocol, physical layer security, privacy amplification, Shannon information.

## I. INTRODUCTION

Solving the key sharing problem between legitimate users, connected by some telecommunication channels, has been in research focus within many years and it is still completely unsolved.

A protocol based on some cryptographic assumption (factoring problem, discrete log problem, error correction algorithm ctr. [1]) has been proposed by Diffie and Hellman [2] many years ago. There are known key distribution protocols based on “key commutative property” of the encryption algorithms [3]. But the corresponding protocol requires to hide the identity of the message sender [4], which is indeed a further cryptographic assumption.

It was developed in recent years a new approach to key distribution problem based on the notion of *physical layer security* (PHY) (see excellent survey [5]). This approach exploits some physical properties of real communication channels connecting legitimate users sharing a secret key in the presence of eavesdroppers. In line with this setting it was published a pioneer paper by A. Wyner [6] and its extension in the papers [7, 8], where legitimate channels were superior to eavesdropper ones on the SNR parameter.

Next, due to advanced Maurer’s papers [9, 10], such approach was extended with the use of so-called *public discussion* and privacy amplification. It enables to transform disadvantage on SNR for legitimate users against eavesdroppers into advantage at the cost of exchange by additional information on public channels.

Other PHY-based protocols execute channels with random parameters (say, fading channels with multipath wave propagation) [9, 10, 11]. And this technique was used also in MIMO-based systems intended for a communication between mobile units [12, 13]. Effective key distribution problem can be solved also in frame of the so-called *quantum cryptography* where special quantum channels and devices [14] should be executed. But it is worth to note that all the key sharing methods mentioned above have been designed for known SNR in the eavesdropper channels or for the case where the number of antennas in the eavesdropper MIMO-based system is limited by some value. However such requirements to enemy system is obviously unrealistic.

Also there is a demand to share secret keys between users connected by constant (practically noiseless) channels (as Internet itself) and without any cryptographic assumption due to a risk of quantum computers to be applied in the future.

In section 2 we remind the key sharing protocol based on extraction of matrix eigenvalues described in [15] as Scheme EVSKey and confirm that it is in fact insecure [16]. Next, we extend this protocol in order to provide the upper bound for SNR in eavesdropper channel. In section 3 we present some channels transform primitives. Section 4 is devoted to results of simulation. In section 5 we optimize protocol parameters to provide both security and reliability of the shared key. Section 6 concludes the paper and proposes some open problem for further investigation.

## II. KEY SHARING PROTOCOL BASED ON EXTRACTION OF MATRIX CHARACTERISTIC POLYNOMIALS

Let us remind the scheme EVSKey [15] used in the current paper in order to generate the binary raw sequence for further creation of the shared key. The scenario corresponding to this scheme is presented in Fig. 1.

Before a transmission, Alice (A) and Bob (B) generate their own reference matrices  $X_A, X_B \in \mathbb{C}^{n \times m}$  with independent matrix elements distributed according to  $CN(0, \sigma_X^2)$  as well as random unitary matrices  $G_A, G_B \in \mathbb{C}^{n \times n}$  where  $n$  is number of antennas employed by each user and  $m$  is the length of pilot signal.

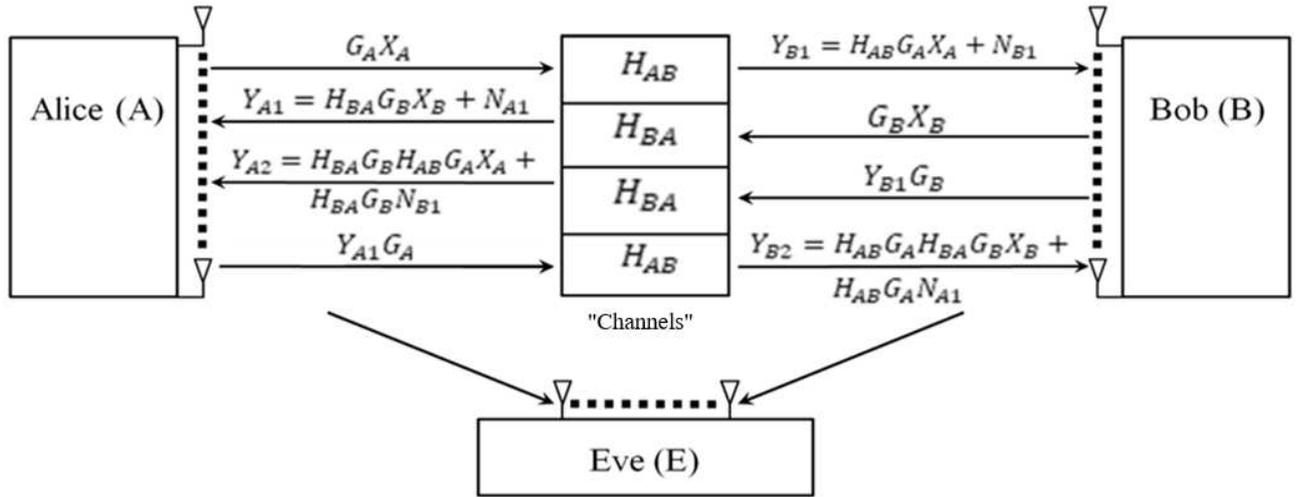


Fig. 1. The scenario corresponding to Scheme EVSKey.

Since in our case A and B are connected by constant noiseless public channels, the original channel matrices  $H_{AB}, H_{BA}$  are generated by A and B as random matrices  $(h_{ABij})_{ij}, (h_{BAij})_{ij} \sim CN(0, \sigma_w^2)$ .  $N_{A1}, N_{B1}$  are AWGN matrices  $(n_{A1ij})_{ij}, (n_{B1ij})_{ij} \sim CN(0, \sigma_e^2)$  generated by users A and B, respectively, as matrices of *artificially created Gaussian noises*. We let further  $\sigma_w^2 = \sigma_x^2 = 1$ ,  $\sigma_e^2 = \sigma^2$ . Let us introduce the following matrices:  $P = H_{BA} G_B$ ,  $Q = H_{AB} G_A$ . Then  $PQ$  and  $QP$  can be estimated by users via the least square method as:

$$PQ = Y_{A2} (X_A)^{-1} \quad (1)$$

$$QP = Y_{B2} (X_B)^{-1} \quad (2)$$

In [15] it was proved that matrices  $PQ$  and  $QP$  have the same non-zero eigenvalues. In [16], it has been proved an extension of such statement, that in fact they have the same *characteristic polynomials* (CP):

$$CP[PQ] = CP[QP] \quad (3)$$

Thus, from (3), we have that the legitimate users A and B are able to extract the same characteristic polynomials after a completion of protocol through noiseless channels although matrices  $PQ$  and  $QP$  can be different. The artificially added noises  $N_{A1}, N_{B1}$  result in errors between shared key bits extracted from quantized characteristic polynomial coefficients, eigenvalues or traces. Therefore, these errors have to be corrected by an additional procedure. Hence, the following question arises – what is the goal of adding artificial noises? The reason is a noising of eavesdropper channel in such a way that power of this noise cannot be decreased by any eavesdropper E!

But firstly we should demonstrate that E is able to intercept even noisy key bits because it was claimed in [15] that it is impossible. Unfortunately the last statement is wrong and in [16] there has been described the procedure about how E is able to intercept key bits, not necessary in the case when she has a close location to legitimate users. In fact, for noiseless channels, if E intercepts  $Y_{A1}, Y_{A2}, Y_{B1}, Y_{B2}$ , where  $Y_{A1} =$

$H_{BA} G_B X_B$ ,  $Y_{A2} = H_{BA} G_B H_{AB} G_A X_A$ ,  $Y_{B1} = H_{AB} G_A X_A$ ,  $Y_{B2} = H_{AB} G_A H_{BA} G_B X_B$ , she can compute the matrix Y:

$$Y = Y_{A2} (Y_{B1})^{-1} Y_{B2} (Y_{A1})^{-1} \quad (4)$$

(We note that pseudo-inverse matrices can be found by Penrose's procedure [17] as

$$(X_p)^{-1} = X^\dagger (X X^\dagger)^{-1}. \quad (5)$$

Here “ $\dagger$ ” is conjugate transpose. It was proved in [16] that matrix Y is *similar* to matrix  $QP$ , thus they have the same characteristic polynomials for nonsingular matrices [18].

Hence the original scheme EVSKey is useless for key sharing but fortunately it can be used as a primary protocol providing lower noisy bound for eavesdropper that cannot be decreased because it is controlled by the legitimate users.

But before we present the following part of key sharing protocol, it is important to show that both artificial noises  $N_{A1}, N_{B1}$  should be added, otherwise eavesdropper can be able to intercept the legitimate key without any errors. Indeed, let us assume that only B creates artificial noise. Then we get:

$$Y_{B1} = Q X_A + N_{B1}, \quad Y_{A2} = P Y_{B1} \\ Y_{A1} = P X_B, \quad Y_{B2} = Q Y_{A1} \quad (6)$$

Next, A extracts CP from the matrix:

$$Y_{A2} X_A^{-1} = P Y_{B1} X_A^{-1} = PQ + P N_{B1} X_A^{-1}, \quad (7)$$

whereas B extracts the key from CP of the matrix:

$$Y_{B2} X_B^{-1} = Q Y_{A1} X_B^{-1} = QP \quad (8)$$

The eavesdropper E extracts the key from CP of the matrix:

$$Y_{A2} (Y_{B1})^{-1} Y_{B2} (Y_{A1})^{-1} = PQ \quad (9)$$

Thus (3) implies that E gets exactly the same key as legitimate user B. This means that such situation has to be excluded.

### III. DESCRIPTION OF CHANNEL TRANSFORM PRIMITIVES

In the following section there will be presented the results of simulation regarding the key bit errors under the provision of two artificial noises  $N_{A1}, N_{B1}$ . If such results give advantage to legitimate users against eavesdroppers, that is  $P_l < P_e$ , where  $P_l, P_e$  are the key *basic bit error rate* (BER) for legitimate users and eavesdropper, respectively, then we

can apply privacy amplification theorem [11]. It states that such algorithm exists which provides an approaching to zero both key BER for legitimate users and Shannon information leaking to eavesdropper with the *key generation rate*:

$$R = h(P_e) - h(P_l), \quad (10)$$

where

$$h(x) = -(x \log_2 x + (1-x) \log_2 (1-x))$$

is the entropy function. But, for opposite situation when it occurs that the key BER's satisfy to inequality  $P_l > P_e$ , it is necessary to apply in advance some additional protocol (primitive) that reduces the previous inequality to opposite one ( $P_l < P_e$ ).

In [11] several examples of such primitives are given. It seems that the best of them is protocol known as "a preference improvement of the main channel" (PIMC). Let us consider the protocol PIMC in more detail, when there are two binary statistically independent symmetric channels without memory (BSC: *binary symmetric channels*): one with BER  $P_l$  and another with BER  $P_e$  and  $P_l > P_e$ . Then legitimate user A has to repeat  $S$  times each bit transmitting over main channel with BER equal to  $P_l$ . Another legitimate user B receives only such  $S$ -blocks which consist of all zeros or ones and takes corresponding decision. He informs over public noiseless channel about blocks that he has accepted and erases other blocks. It is easy to see that such protocol forms the following BER for B:

$$\tilde{P}_l = \frac{P_l^S}{P_l^S + (1-P_l)^S} \quad (11)$$

At the same time eavesdropper E intercepts  $S$ -blocks over BSC with BER  $P_e$  and controls public noiseless channels. E knows exactly which  $S$ -blocks are accepted by B. But because E's channel is statistically independent with the main channel (A→B), she should take decision about bits corresponding to  $S$ -block using *majority rule*. This means that she takes a decision that  $S$ -block carries bit "0", if this block has more zeros than ones and decision about bit "1", if the number of ones in that  $S$ -block is larger than the number of zeros. Then the BER after such decision will be for odd  $S$  the following:

$$\tilde{P}_e = \sum_{i=\frac{S+1}{2}}^S \binom{S}{i} P_e^i (1-P_e)^{S-i} \quad (12)$$

But unfortunately, it seems to be impossible to repeat bits if they were extracted from CP's of the matrices PQ and QP!

In order to avoid this problem let us modify slightly our previous protocol as it is shown in Fig. 2. We can see that just after a generation of "raw" bits from matrices PQ and QP, user B generates truly random binary string  $\gamma$  that is XOR-ed with B's raw bits  $K_B$  and it is transmitted over public and noiseless channel to user A that adds this string with her raw bits  $K_A$  in order to get:

$$\tilde{K}_A = K_B \oplus \gamma \oplus K_A = K_A \oplus \varepsilon_{AB} \oplus K_A \oplus \gamma = \gamma \oplus \varepsilon_{AB}, \quad (13)$$

where  $\varepsilon_{AB}$  is discrete noise string between raw key strings  $K_A$  and  $K_B$ . It is easy to see that in such setting the user B is able already to repeat  $S$ -times each bit of  $\gamma$  in order to perform the previous protocol. From now on we consider just  $\gamma$  as a new key string, transmitted to A over BSC with BER equal to  $P_l$ .

At the same time E, having received  $K_B \oplus \gamma$  and her raw key  $K_e$ , extracted by (4), sums these sequences up. This gives:

$$\tilde{K}_e = K_e \oplus \gamma \oplus K_B = K_B \oplus \varepsilon_{BE} \oplus \gamma \oplus K_B = \gamma \oplus \varepsilon_{BE}, \quad (14)$$

where  $\varepsilon_{BE}$  is discrete noise string between raw key strings  $K_B$  and  $K_e$ , that is equivalently to a transmission of key string  $\gamma$  to eavesdropper E over BSC with BER equal to  $\tilde{P}_e$ .

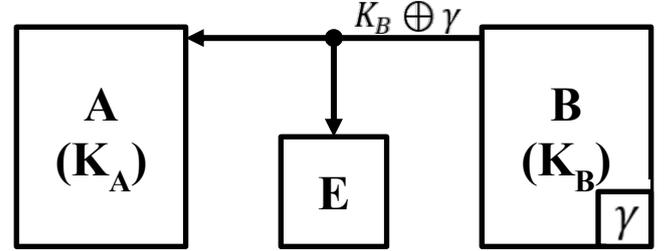


Fig. 2. Modified key sharing protocol.

#### IV. RESULTS OF SIMULATION

##### A. Using quantized matrix traces as the raw key bits

Since the traces of matrices are complex, they can be quantized both on amplitude and on phase. It was proved in [16] that the quantized intervals on amplitude of the traces providing equal probabilities of their occurrence should be chosen as follows:

$$r_{k-1} \leq |Z| < r_k, \quad k = 1, 2, \dots, N, \quad (15)$$

where  $Z$  is the trace of the matrices,  $r_k = \sigma_z \sqrt{-\ln(1 - \frac{k}{N})}$ ,  $\sigma_z^2 = n^2 \sigma_w^2 (\sigma_w^2 + \sigma_e^2) + n \sigma_e^2$ ,  $N$  is the number of intervals. In table 1 there are presented the results of BER simulation for  $N = 16$ , different values of NSR, and different matrix sizes  $n \times m$ . We see from this Table that for all parameters  $P_l > P_e$  and hence it is necessary to execute the protocol PIMC (see section III) in order to reduce to opposite situation  $P_l < P_e$ , that will be demonstrated in the sequel.

##### B. Using quantized matrix eigenvalues as the raw key bits

Unfortunately, there appears one problem in this case – how to compare the numbering of eigenvalues adopted by different users? Let us denote by  $N_p$ ,  $N_A$  the numbers of quantization intervals on phase and on amplitude respectively.

TABLE 1.

SIMULATION RESULTS OF THE BER FOR EXTRACTION THEM FROM MATRICES TRACES BOTH LEGAL USERS ( $P_l$ ) AND EAVESDROPPER ( $P_e$ ) WITH 8 SECTORS AND 8 RINGS, UNIFORM PHASE QUANTIZATION AND AMPLITUDE STEP QUANTIZATION BY (15)

$\sigma^2$ \ n/m	4x4	4x6	4x12	8x8	8x16	16x16
	$P_l, P_e$					
0.1	0.348 0.274	0.255 0.191	0.155 0.116	0.363 0.291	0.152 0.118	0.364 0.303
0.01	0.212 0.157	0.104 0.075	0.058 0.044	0.209 0.139	0.055 0.043	0.219 0.158
0.001	0.098 0.063	0.032 0.022	0.013 0.011	0.085 0.063	0.015 0.012	0.098 0.064

Let  $N = N_p \times N_A$  be total number of quantization intervals. Then we find the number of eigenvalues that hits each of the  $N$  interval (cells). After a completion of eigenvalues extraction, we get a string of integers  $g_1, g_2, \dots, g_i, \dots$ , where  $g_i$  is the number of the  $i$ -th cell containing at least one eigenvalue. If several eigenvalues occur in the same cell, then the cell number is repeated as  $g_i, g_i, \dots$ . Next each number  $g_i$  is presented as a bit string and such strings are connected in a consecutive binary manner. The final binary string forms the raw shared key. It is easy to see that the total number of bits for each session of protocol can be computed as [16]

$$\log_2 \binom{N+n-1}{n} = \log_2 \frac{(N+n-1)(N+n-2)\dots N}{n!} \quad (16)$$

In Table 2 there are presented the results of BER simulation for different matrix sizes and different NSR for eigenvalues extracted from matrices where each eigenvalue is quantized on 8 sectors and 8 rings.

TABLE 2.

SIMULATION RESULTS OF THE BER FOR EXTRACTION OF THEM FROM MATRIX EIGENVALUES BOTH LEGAL USERS ( $P_l$ ) AND EAVESDROPPER ( $P_e$ ), WITH 8 SECTORS AND 8 RINGS FOR EACH EIGENVALUE

$\sigma^2$	n/m	4x4	4x6	4x12	8x8	8x16	16x16
		$P_l, P_e$					
0.1		0.348	0.262	0.170	0.350	0.207	0.207
		0.288	0.204	0.121	0.302	0.159	0.159
0.01		0.215	0.115	0.069	0.235	0.085	0.085
		0.156	0.080	0.049	0.175	0.057	0.057
0.001		0.104	0.037	0.022	0.127	0.029	0.029
		0.068	0.027	0.014	0.082	0.021	0.021

We see from this Table also that, as before, for all BER parameters,  $P_l > P_e$ , hence it is necessary to execute the protocol PIMC in order to provide the opposite situation ( $\tilde{P}_l < \tilde{P}_e$ ). We show in the sequel how to do it.

#### V. OPTIMIZATION OF KEY-SHARING PROTOCOL PARAMETERS IN ORDER TO PROVIDE GIVEN SECURITY AND RELIABILITY

It has been proved by the *Enhanced Privacy Amplification Theorem* [19], that the eavesdropper's expected Shannon information  $I_o$  about the final key sequence shared by legitimate users, satisfies the inequality:

$$I_o \leq \frac{2^{-(k-t_c-l_0-r)}}{\alpha \ln 2}, \quad (17)$$

where  $k$  is the length of the string  $x$  generated by A and B after a completion of the protocol PIMC,  $t_c$  is the Renyi (or collision) information obtained by eavesdropper E about the string  $x$  received by E through a BSC with BER equal to  $\tilde{P}_e$ ,  $r$  is the number of check bits sent by one of legitimate users to another one in order to reconcile their string,  $l_0$  is the length of the final key,  $\alpha$  is a coefficient that approaches to 0.42 for any fixed  $r$ , as  $k$ ,  $r$  and  $k-r$  are increasing (we recall that the *privacy amplification procedure*, providing the inequality (17), can be performed in two stages: firstly with the use of a hash function chosen randomly from universal<sub>2</sub> class and, secondly, by special "puncturing" of hash string [19]).

Let us consider a scenario, that allows to optimize parameters:  $k, r, S$  (see (11), (12)) for given prior values  $l_0, I_o$

and  $\tilde{P}_{ed}$  – the probability of incorrect decoding of final key string.

1. Given  $I_o$ , find the bound value

$$k - t_c - l_0 - r = -\log_2(I_o \alpha \ln 2) = \lambda_1 \quad (18)$$

2. Calculate the value Renyi entropy [19]:

$$H_c = -\log_2(\tilde{P}_e^2 + (1 - \tilde{P}_e)^2) \quad (19)$$

3. Taking into account the relation

$$t_c = k - kH_c, \quad (20)$$

we get by (18)

$$kH_c - r = \lambda_1 + l_0. \quad (21)$$

4. In order to provide a decreasing of  $\tilde{P}_{ed}$  for bit string of length  $k$  and with execution of  $r$  check bits it is necessary to satisfy Shannon's inequality [21]:

$$\frac{k}{k+r} < C, \quad (22)$$

where

$$C = 1 + \tilde{P}_l \log_2 \tilde{P}_l + (1 - \tilde{P}_l) \log_2 (1 - \tilde{P}_l) \quad (23)$$

5. Substituting (19) into (20) and considering (21) jointly with (22) (taken as equality) it is possible to solve the linear system of equations with respect to  $k$  and  $r$ .

We can take different values  $P_l, P_e$  from simulation results (see Tables 1, 2) and, by varying the parameter  $S$  into (11), (12), to obtain the new values  $\tilde{P}_l, \tilde{P}_e$ , that would improve our protocol. For example one could increase the length of final key  $l_0$  or to make it more secure by decreasing the value  $I_o$ . It is worth to note that we do not find so far a final key reliability in terms of the value  $\tilde{P}_{ed}$  but *we only guaranty* (due to Shannon's theorem) the existence of such encoding and decoding procedures that provide an approaching of this probability to zero.

Selection of the constructive encoding/decoding procedures requires further research. Seemingly, it should be of well known class of codes like LDPC. The later approaches *the Shannon limit* for large block lengths [20]. But before we face with some examples, it is necessary to fix the value  $I_o$  by a reasonable manner. Let us present a lower bound for  $\tilde{P}_{ed}$  based on Fano's inequality [21]:

$$H(U/V) \leq h(\tilde{P}_{ed}) + \tilde{P}_{ed} \log_2(M-1), \quad (24)$$

where  $H(U/V)$  is *conditional entropy* for eavesdropper E;

$$h(x) = -x \log_2 x - (1-x) \log_2 (1-x), 0 \leq x \leq 1 \quad (25)$$

$M$  is the number of possible keys (in our case it is equal to  $2^{l_0}$ );  $\tilde{P}_{ed}$  is the probability of incorrect decoding that means a transition of the key string to another one (it is worth to note that the meaning of inequality (24) is the following: if entropy  $H(U/V)$  is large, then the probability  $\tilde{P}_{ed}$  of incorrect decoding cannot be small). The graph of the function  $\mu(\tilde{P}_{ed}) = h(\tilde{P}_{ed}) + \tilde{P}_{ed} \log_2(M-1)$  is shown in Fig. 3.

We can see from Fig. 3 that if  $H(U/V)$  is larger than some value, say  $H_0$ , then  $\tilde{P}_{ed}$  should be at least  $P_{ed}^0$  (see Fig. 3). Thus for given  $M = 2^{l_0}$  and  $I_o$ , we can find the lower bound

for  $\widetilde{P}_{ed}$  and if it occurs very close to the value  $(M-1)/M$  (the probability of a *random key string guessing*) then it is assumed that the key sharing protocol is secure. If we let  $I_0 = 10^{-3}$ ,  $M = 2^{64}$ , then  $H(U/V) = 64 - 0.001 = 63.999$ . Using the graph of  $\mu(\widetilde{P}_{ed})$  we get that  $\widetilde{P}_{ed}$  is sufficiently close to the case of random guessing  $(M-1)/M$ .

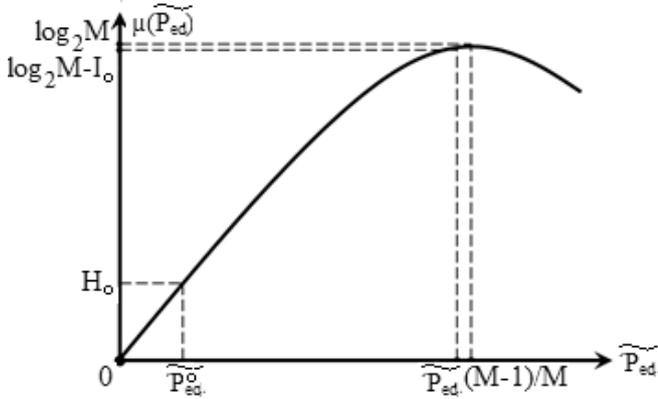


Fig. 3. Graph of function  $\mu(\widetilde{P}_{ed})$  against  $\widetilde{P}_{ed}$ .

*Examples:*

1. Let us take from Table 2 the parameters  $n \times m = 4 \times 4$ ,  $\sigma^2 = 0.1$ . Then  $P_e = 0.29$ ,  $P_l = 0.35$ . Select  $S = 5$  in (12), (13). Then we get by (12, 13)  $\widetilde{P}_l = 0.043$ ,  $\widetilde{P}_e = 0.15$ ,  $H_c = 0.425$  and  $C = 0.439$  by (20) and (24). Selecting  $\lambda_1 = 10$ ,  $l_0 = 64$  and following to the scenario steps 1÷5, we get finally for key size 64 bit  $k = 1058$ ,  $r = 374$ ,  $I_0 \leq 2^{-10} \approx 10^{-3}$ .

2. Let us take the same as in Example 1 initial parameters  $P_l$ ,  $P_e$  and the same  $S = 5$ . But let us increase  $\lambda_1$  till 30. Then we get finally  $k = 1337$ ,  $r = 472$ ,  $I_0 \leq 2^{-30} \approx 10^{-9}$ .

So we can see that it is possible to provide better security by changing protocol parameters. Because in this case the inequality (23) coincides with equality, it is necessary to decrease slightly the parameter  $k$  in order to provide approaching of  $P_{ed}$  to zero by Shannon theorem.

3. Let us increase the key size  $l_0$  up to 128, because the most of contemporary encryption standards (like GOSI-2015 and AES) have namely such key sizes. We assume the same initial probabilities  $P_l$ ,  $P_e$  as before and the same  $S = 5$ . Following to scenario steps 1÷5 we get the parameters:  $k = 2228$ ,  $r = 787$ .  $I_0 \leq 2^{-30} \approx 10^{-9}$ . We can see from this example that it is possible to share more longer key with a good security at the cost more longer error correcting code.

4. In this example we consider the case of key bit extraction from matrix traces (see Table 1).

Let us select the parameters  $n \times m = 4 \times 4$ ,  $\sigma^2 = 0.1$ . Then we can see from Table 1 that  $P_l = 0.348$ ,  $P_e = 0.274$ . Selecting parameter  $S = 7$ , we get by (12), (13) that  $\widetilde{P}_l = 0.012$ ,  $\widetilde{P}_e = 0.095$ . Following to scenario steps 1÷5 we compute for  $l_0 = 128$ , that  $k = 962$ ,  $r = 101$ ,  $I_0 \approx 10^{-9}$ .

We can see that having selected such protocol parameters  $n \times m$  and  $NSR = \sigma^2$ , we can perform a tradeoff between

security ( $I_0$ ), reliability ( $P_{ed}$ ) and error correction procedure complexity that is proportional to  $k$  and  $r$ .

In Fig. 4 there is presented a diagram of all procedures that must be executed in order to complete the key sharing protocol among legitimate users connected by noiseless, public and constant parameter communication channel. There is a new block (verification of key string authenticity) that has not been discussed before. In fact, this procedure is requested for any key sharing protocol in presence of an active adversary (eavesdropper). Otherwise the adversary can impersonate legitimate users and eventually share with them common key. It is common to use authentication method based on the so-called *short-key* [22]. The Needham-Schroder authentication protocol [23] can be used if users have initially distributed, by some trusted center, short keys.

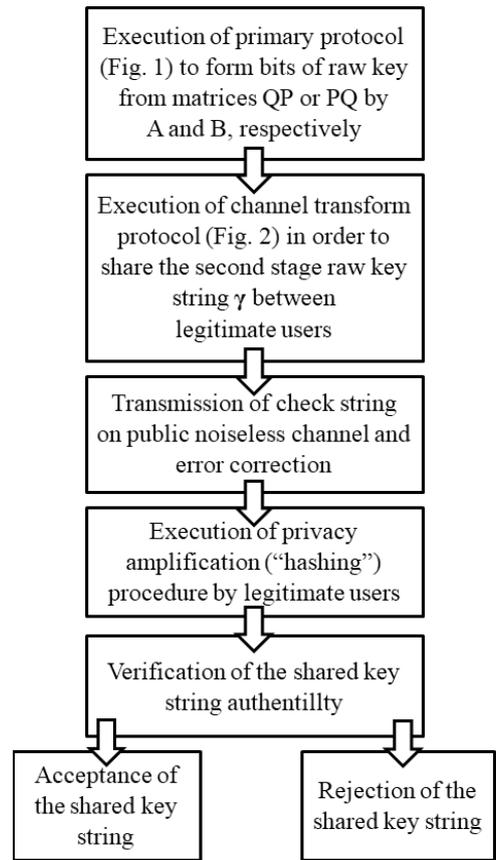


Fig. 4. Diagram of the whole key sharing protocol.

Another way is if users can provide the so-called *paring procedure* during their “face to face” meeting (like Mag Pairing or Physical vibration [24, 25]).

It is interesting to estimate (at least roughly) the length of the whole protocol (in transmitting channel bytes). Our computations show that the whole key sharing protocol requires about 100 Kbytes channel uses to produce 128 key bits.

## VI. CONCLUSION

We have proposed key sharing protocol for noiseless public constant parameter communication channels (like Internet or “Direct seen”). The main novelty of our scenario is that *it is not based on some unrealistic assumptions* like given SNR, cryptographic assumption for eavesdropper (hard factoring problem) or multipath wave propagation, that is different for legitimate users and eavesdropper. The core of our protocol is the *Scheme EVSKey* proposed in [15]. But we proved that such protocol itself is insecure. Therefore we modified it by introducing artificial noise by legitimate users that does not allow to decrease this noise power by eavesdropper. Next we apply effective procedure of privacy amplification that provides both security and reliability for legitimate users. It is worth to note that good statistical properties of the final key string follow directly from such properties of truly random generated  $\gamma$  (see Fig. 2). It seems at a first glance that the paper [26] was devoted also to a solution of the same problem as our paper. In fact, it has only one common notion – “artificial noise”, but many differences, namely:

- we consider key sharing problem, instead of secure information transmission as in [26],
- in [26] it is executed either a MIMO system in fading channels or a set of “helpers”; our protocol is used in constant parameter public channel due to information exchange between two users,
- in [26] it is created noise in “zero-space”, whereas we execute special protocol imposing to eavesdropper artificial noise,
- in [26] it is provided zero noise by “zero-forcing”, but we provide a lower bound only for noise power,
- finally, in [26] it is guaranteed only some given *secrecy capacity*, but it is unknown how to realize it, namely how to provide constructive encoding/decoding procedures? But we on the contrary calculate Shannon information leakage to eavesdropper after application of the known privacy amplification procedure and find the parameters  $n$  and  $k$  for linear error correcting codes. Next investigations in the direction of artificial noise can be found in [27, 28].

The problems for further investigation are:

- consideration of constructive error correction procedures and
- elaboration of effective authentication algorithm against an active adversary.

## REFERENCES

- [1] A. J. Menezes, P. C. van Oorschot, and S. A. Vanstone, Handbook of Applied Cryptography, ser. The CRC Press series on discrete mathematics and its applications. 2000 N.W. Corporate Blvd., Boca Raton, FL 33431-9868, USA; CRC Press, 1997. ISBN 0-8493-8523-7
- [2] W. Diffie and M. E. Hellman, “New directions in cryptography,” vol. 22, no. 6, pp. 644–654, 1976.
- [3] Schneier B., “Applied Cryptography”, JW Incomp., 1994.
- [4] B. Alpern and F. B. Schneider, “Key exchange using ‘keyless cryptography’.” Inf. Process. Lett., vol. 16, no. 2, pp. 79–81, 1983. [Online]. Available: <http://dblp.unitrier.de/db/journals/ipl/ipl16.html#AlpernS83>
- [5] A. Mukherjee, et al. “Principles of Layer Security in Multiuser Wireless Network: A Survey”, arXiv:1011.3754.3 [cs. IP], 2014.
- [6] A. Wyner, “Wire-tap channel concept,” Bell System Technical Journal, vol. 54, pp. 1355–1387, 1975.
- [7] I. Csiszár and J. Körner, “Broadcast channel with confidential messages.” IEEE Transactions on Information Theory, vol. 24, no. 2, pp. 339–348, 1978.
- [8] V. Korjik and V. Yakovlev, “Non-asymptotic estimates for efficiency of code jamming in a wire-tap channel,” Problems of Information Transmission, vol. 17, pp. 223–222, 1981.
- [9] L. H. Ozarow and A. D. Wyner, “Wire-tap channel II,” in Advances in Cryptology: Proceedings of EUROCRYPT 84, A Workshop on the Theory and Application of Cryptographic Techniques, Paris, France, April 9–11, 1984, Proceedings, 1984. doi: 10.1007/3-540-39757-4\_5 pp. 33–50. [Online]. Available: [https://doi.org/10.1007/3-540-39757-4\\_5](https://doi.org/10.1007/3-540-39757-4_5)
- [10] V. Korjik and D. Kushnir, “Key sharing based on the wire-tap channel type ii concept with noisy main channel,” in Proc. Asiacrypt96. Springer Lecture Notes in Computer Science 1163, 1996, pp. 210–217.
- [11] U. Maurer, “Secret key agreement by public discussion from common information.” IEEE Transactions on Information Theory, vol. 39, no. 3, pp.733–742, 1993.
- [12] V. Yakovlev, V. I. Korzhik, G. Morales-Luna, “Key distribution protocols based on noisy channels in presence of an active adversary: Conventional and new versions with parameter optimization,” IEEE Transactions on Information Theory, vol. 54, no. 6, pp. 2535–2549, 2008.
- [13] V. Korjik and M. Bakin, “Information-theoretically secure keyless authentication,” in Proc. IEEE Symp. on IT’2000. IEEE, 2000, p. 20.
- [14] C. H. Bennett, F. Bessette, G. Brassard, L. Salvail, and J. Smolin. “Experimental quantum cryptography”, J. Cryptol., vol. 5, no. 1, pp. 3–28, Jan. 1992. [Online]. Available: <http://dl.acm.org/citation.cfm?id=146395.146396>
- [15] D. Qin and Z. Ding, “Exploiting multi-antenna non-reciprocal channels for shared secret key generation,” IEEE Transactions on Information Forensics and Security, vol. 11, no. 12, pp. 2693–2705, Dec 2016. doi: 10.1109/TIFS.2016.2594143
- [16] Starostin V.S. et al “Key Generation protocol executing through non-reciprocal fading channels”, Intern. Journal of Computer Science and Applications, vol. 16, no. 1, pp. 1–16, 2019.
- [17] Ben-Israel, Adi; Greville, Thomas N.E. , p. 7. Generalized inverses: theory and applications (2nded.). NY: Springer. ISBN0-387-00293-6, 2003.
- [18] Home and Johnson, “Matrix Analysis”, Cambr. Univ.Pres. 1985.
- [19] V. Korjik, G. Morales-Luna, and V. Balakirsky, “Privacy amplification theorem for noisy main channel,” Lecture Notes in Computer Science, vol. 2200, pp. 18–26, 2001.
- [20] K. Shalkoska, Implementation of LDPC Algorithm: In C Programming Language. LAP LAMBERT Academic Publishing, 2017. ISBN9783330026049. [Online]. Available: <https://books.google.com.mx/books?id=1yNcMQAACAAJ>
- [21] Fano R.M. Transmission of Information. A statistical theory of communication, Willy Bullisher, 1961.
- [22] D. Dasgupta, A. Roy, and A. Nag, Advances in User Authentication, 1st ed. Springer Publishing Company, Incorporated, 2017. ISBN 3319588060,9783319588063
- [23] R.M. Needham and M.D. Schroeder, “Using Encryption for authentication in Large Network of computers”. ACM, v21, p.993-999, 1978.
- [24] Jin R. et al “ MagPairing: Pairing Smartphones in close proximity using magnetometer”, IEEE Trans. of Information Forensics and Security, 6, p. 1304-1319, 2016.
- [25] Roy N. et al, “Faster Communication through Physical vibration”, proc USENIX Symp. Netw. Syst. Design, p. 671-675, 2016.
- [26] Goel S. and Negi R., “Guaranteeing Secrecy using Artificial Noise”, IEEE Trans. of Wireless Communication, vol. 7, no 6, p. 2180-189, 2008.
- [27] Bangwon Seo, “Artificial Noise Based Secure Transmission Scheme in Multiple Antenna Systems”, International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 21 (2016)
- [28] Liu, S., Hong, Y., & Viterbo, E. Artificial noise revisited. *IEEE Transactions on Information Theory*, 61(7), 3901 - 3911. <https://doi.org/10.1109/TIT.2015>.

# License Plate Detection with Machine Learning Without Using Number Recognition

Kazuo Ohzeki

Algorithm Lab.

Nishida Build. 5F 2-14-6 Shibuya,  
Shibuya-ku 150-0002 Tokyo Japan  
Email: ohzeki@shibaura-it.ac.jp

Max Geigis

University of Applied Science

Bahnhofstraße 61, 87435  
Kempten (Allgäu) Germany  
Email: maxgeigis@gmail.com

Stefan Alexander Schneider

University of Applied Science

Bahnhofstraße 61,  
87435 Kempten (Allgäu) Germany  
Email:  
stefan-alexander.schneider@hs-kempten.de

**Abstract**—In autonomous driving, detecting vehicles together with their parts, such as a license plate is important. Many methods with using deep learning detect the license plate based on number recognition. However, there is an idea that the method using deep learning is difficult to use for autonomous driving because of the complexity in realizing deterministic verification. Therefore, development of a method that does not use deep learning (DL) has become important again. Although the authors have made the world's best performance in 2018 for Caltech data with using DL, this concept has now turned to another research without using DL. The CT5L method is the latest type, that includes techniques of the continuity of vertical and horizontal black-and-white pixel values inside the plate, unique Hough transform, only vertical and horizontal lines are detected, the top five in the order of the number of votes to ensure good performance. In this paper, a method to determine the threshold value for binarizing input by machine learning is proposed, and good results are obtained. The detection rate is improved by about 20 points in percent as compared to the fixed case. It achieves the best performance among the conventional fixed threshold method, Otsu's method, and the conventional method of JavaANPR.

## I. INTRODUCTION

THE issue of practical use of automatic driving is discussed in ROAD2016 [1]. With the advancement of technology, the number of test cases is said to be huge and to be  $10^{12}$ . Because of this, Virtual Test becomes more important. With the addition of AI technology, this type of test case is expected to become even more extensive, and it is also pointed out that the deterministic test becomes more difficult.

With the technology that processes images obtained from camera sensors and recognizes them in the external world, many methods can be developed and studied by software development on a computer. Our group has been promoting the Samurai Project on ADAS since 2016, and has developed image processing technologies such as license plate detection, lane detection, vehicle detection, vehicle maker logo detection, front grill detection, etc. . Among them, LPD by deep learning has developed an improved method that surpasses the highest performance of the past [2]. Currently, many LPDs detect plate areas by number and character recognition.

Among the technologies related to autonomous driving, we are especially developing technologies for vehicle detection and license plate detection. In addition to vehicle detection, the significance of plate detection will be described. As a second step after vehicle detection, for example, plates, lights, etc. of the vehicle elements are individually detected. The verification of the element by this detection makes it possible to verify whether the event that detected the vehicle was correct. This can increase the reliability of vehicle detection. In the vehicle detection, as a detection result, both a correct detection of a vehicle (True Positive: TP) and a false detection (False Positive: FP) in which a non-vehicle is recognized as a vehicle are output. That is, detection of precision =  $TP / (TP + FP)$  is performed. In this equation, if the detection result of the vehicle element such as plate detection is evaluated as the reliability in the second stage processing and the false detection (FP) is eliminated, the FP value of the equation can be reduced and the precision rises. This is the significance of performing license plate detection in addition to vehicle detection.

Deep learning is also effective in license plate detection [2]. Convolutional Neural Network (CNN), which is used in deep learning, performs effective computations, but it is difficult to clearly describe the process of generating the correct result by including the multiplication of many coefficients and non-linearity. Therefore, there is a high possibility that the test of the autonomous driving vehicle using deep learning can not estimate the total number within the feasible range, which is a problem [3]. It has also been pointed out that there are problems with AI and safety assurance by the Cyber-Physical System research group and the researchers of software engineering. [4] Therefore, developing a scheme that does not use CNN is effective as one method to avoid the risk arising from such an indication.

Estimated power consumption of PCs in electric vehicles is 40% of the total increase [5], and a decrease is required. Low power consumption is desired. Methods of non-deep learning etc. are also becoming more important as a study target.

Another background is that the illegality of ANPR was pointed out in 2008 [6] [7]. The use for the crime prevention such as the police is established as a legitimate use even after that. However, there is a move that appeals to identify personal information from the number as a privacy issue, and it remains unsolved, and there is a possibility that the future controversy will continue. It is unpredictable how future laws will change, and insurance technology development and development of backup technology are necessary to cope with the uncertainty. Therefore, in consideration of the privacy problem, we have taken measures to detect the plate only from the features of the license plate without using number recognition.

In this paper, considering such background, it is characterized by examining in the range of methods that do not use number recognition and do not use deep learning such as CNN.

In our group, we are conducting research and development on practical application of automatic driving in the Samurai Project. Regarding license plate detection, Samurai Project has developed both statistical and deterministic methods, as shown in Table 1. Although Linear Regression used to determine the threshold is due to machine learning, it is a fixed factor, a finite number of realizable deterministic processes, and not deep learning. The reason that deep learning can only be performed by statistical test is considered to be due to the large number of coefficients and the fact that the circuit contains nonlinearity. Along with the development of test environment in the future, in addition to pursuing high performance using deep learning, it is also necessary to carry out technological development without deep learning and secure flexibility in technology integration .

Table 1 Two alternatives of Samurai Project for plate detection methods

item	method1	method2
process/ test	statistic non-linear dis-continuous	deterministic linear continuous
algorithm	deep Learning CNN	CT5L (threshold binarization Hough transform machine Learning Linear regression)
reference	[2]	[19], [20] [This paper]

Further structure of this paper is as follows. Section II provides the overview of the relevant works. Section III provides our number plate detection algorithm. Section IV describes machine learning tools used in this paper. Section V demonstrates the experimental results. Evaluation, Conclusions and future research directions are outlined in section VI.

## II. RELATED WORKS

### A. Methods by Conventional Image Processing

The license plate detection of a vehicle includes a method by conventional image processing and a method by recent deep learning, which are respectively useful techniques. First, as a conventional method of image processing, there is a method using Rectangle Detection by Rosito et al. [8]. They made it possible to detect a generalized rectangle including the case of tilting using the rectangle and symmetry of the rectangle in the Hough transform. This is used to detect License Plate and Rectangle-like semiconductor chips.

There is a scheme of Frequency of Luminance changes in the Vertical and horizontal direction, which has been studied by Martinsky [9]. The surface of the vehicle is flat, and in the area of the License Plate, there are large brightness changes in numbers, letters, and the like. If scanning horizontally, the place where the change is large corresponds to the plate area. Also, if scanning vertically, the plate part will be the place of large change as well.

### B. Methods Using Deep Learning

Next, we describe a high-performance method using deep learning. Zang et al use a CNN for three datasets of RGB color channels, and integrates the results by majority decision [10]. They show the advantage of three channel integration. Montazzolli et al. detect Brazilian plates using YOLO in real-time [11]. There are three stages, cropping small region around plate, detecting plate region, and character recognition on the plate by adjusting parameters of YOLO.

Dong et al. also present two-stage detection of Chinese plates [12]. At the first stage, it detects plate region using Region Proposal Network in low resolution, then replaces by a high resolution image, and detects four vertices of the plate using Faster R-CNN, then obtains corrected plate region using affine transform. At the second stage, seven STN and CNN units work in parallel for seven characters for separation and recognition.

The highest performance using Deep learning for standard dataset of Caltech is presented by Kim et al. [13], together with the search from the second to the seventh results. They use the faster R-CNN for the vehicle region detection and candidates for license plates in each detected region with the hierarchical sampling method (CNN) are generated. Finally, non-plate candidates are filtered out by training a deep convolutional neural network. Training two different CNN's for plates and non-plates, they remove FP results using non-plate CNN. For Caltech standard dataset, precision of 98.39% and recall of 96.83% are performed, which are the best world records at the time of publication 2017. But the method by our group, which was announced in

2018, surpasses the method of Kim et al. Achieved performance [2].

By improving accuracy of character recognition, a method to detect character region at the first stage without vehicle region detection has become effective. Among them, [14] is an ambitious paper using deep learning of character recognition. At the first stage, from candidate region detected by weak character detection making saliency map, rectangle plate region is detected after removing FP by two-class CNN. At the second stage, using character separation and character recognition, together with labeling results based on connectivity, numbers and characters on the plate are confirmed. In total 37 class CNN is constructed with ten numbers and 26 uppercase letters and a single non character. Region detection at the first stage should be improved, while character recognition at the second stage improves by adding connectivity process. Using ten numbers and 26 uppercase letters for CNN, the recognition rate can be advanced, though more kinds of characters are needed depending on each country specification. For the first stage of region detection of a plate, character recognition may fail to produce FP's for logos and advertisements other than plates. In the case of serial construction, total performance is a product of each performance of each stage. Accuracy of each stage must be the highest. Even in the case of so-called Coarse-to-Fine serial construction, the first stage of Coarse should not be coarse but fine in accuracy, excluding FP and removing FN.

ANPR software vendors have published Accuracy results based on image benchmarks. In 2017, Sighthound announced an accuracy of 93.6% for original images [15]. In 2017, OpenALPR's commercial software announced 95-98% accuracy rates in public images [16]. The Brazilian team announced a scheme with an average recognition rate of 93.53% in April 2018, showing a significant improvement from the 81.8% obtained in the previous paper [17].

In the paper by Silva et al. at the 2018 ECCV, an example of detection in a moving picture is shown, but in the comparison of Accuracy, the result changes depending on the data set used [18]. As quoted in Table 2-1, the data averages 81-89% with the lowest at 57-75% and the highest at 96-98%.

Looking at these results, LPD for autonomous operation is still in development, and future performance improvement is expected. In deep learning, it is expected that the amount of computation is large and thus the power consumption will also increase. According to materials of Michigan University [5], 40% of the power consumption in the vehicle is the processing of PC, and the reduction is an issue. Development of LPD by conventional image processing technology not using deep

learning adapted to such a requirement is required in the menu of the autonomous driving system. Further, a method using a technique for recognizing numbers and characters on a plate as a key is in conflict with German precedents, and a method for performing feature detection of a plate that does not perform number or character recognition is required. For these reasons, in this paper, we will develop LPD by Image Processing.

### III. PROPOSED METHOD CT5L

In addition to vehicle detection, the significance of plate detection is to individually identify vehicle elements such as plates, lights, radiators, logos, tires, window glasses, shadows, tail lamps, etc., as a second step after vehicle detection. Thus, it is possible to verify whether the event that detected the vehicle was correct. This can increase the reliability of vehicle detection (see Fig. 1). Vehicle detection includes, as detection results, both correct vehicle detection (True Positive: TP) and false detection (False Positive: FP) in which a non-vehicle is recognized as a vehicle. That is, precision =  $TP / (TP + FP)$ . In this equation, detection results of vehicle elements such as plate detection are evaluated as reliability in the second stage processing. Specifically, if detection of each element is confirmed, one point is added, and if the sum becomes larger than  $n$ , it is regarded as correct detection, otherwise it is regarded that vehicle detection is erroneous. Eliminating false positives (FPs) can reduce the FP value of the equation and

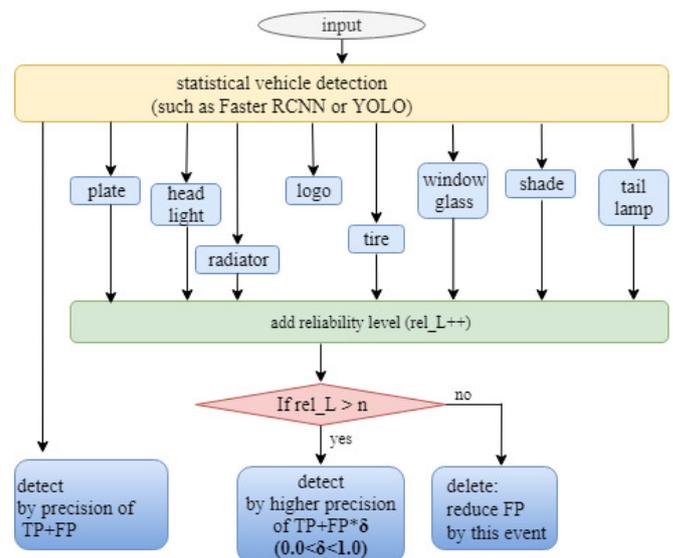


Fig.1 Integration of vehicle and element detection

raise the precision value. This is the significance of performing license plate detection in addition to vehicle detection.

We will describe the "Combining Top Five Lines (CT5L) Method" of the proposed method. This method is an improvement of the basic method presented in [19] and [20]. In the process of development, there were several branched versions, such as the oblique detection compensation method by Okunuki and the 3D compensation method by Max Geigis. In this paper, based on the examination of variations of this, we constructed a high-performance and stable method. Fig. 2 shows a block diagram of CT5L Method.

The number plate area determined from the vehicle area of the input image is compared with the threshold value and binarized (binary image). The binarized data is evaluated as to vertical change and continuation length, horizontal change and continuation length as processing before Hough transform. Based on this evaluation result, points that are candidates for plate boundaries are registered, and points that are not so are deleted (VH continuous condition).

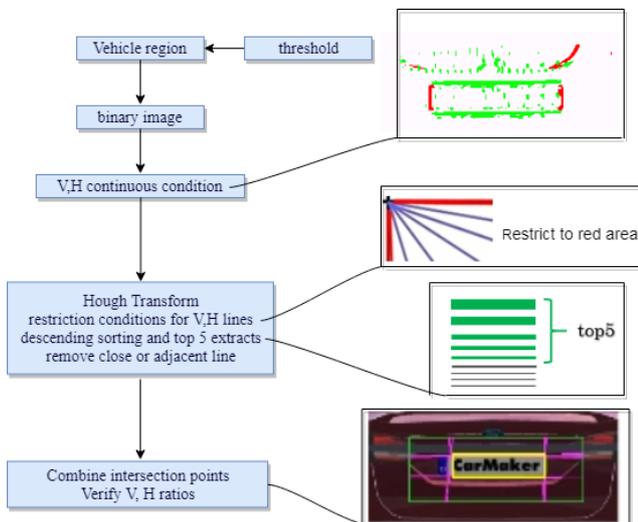


Fig.2 CT5L Method Block Diagram.

Next, Hough transform is performed in the "Hough Transform" section. The Hough transform is performed only within a range in which the angles are limited to  $0 \pm \epsilon$  degrees and  $270 \pm \epsilon$  degrees. The transform is performed only on the red area shown in Fig. 2. As a result, unnecessary calculations are not performed, and thus speeding up can be achieved. When the search range is limited to plus and minus 3 degrees, the amount of operation is reduced to  $12/360 = 1/30$ . Sorting is performed in descending order of the number of votes of distribution after transform, and the upper five lines are selected for each of vertical and horizontal. At this time, close lines may be dense, so if the absolute value of the difference between the line detected later and the upper-rank line up to that time is less than  $\delta$ , it

will be regarded as the same line and removed from the ranking process. In this way, five valid lines are extracted as candidates for the plate outline. Next, in the "Combination of intersection point", select a pair that forms a rectangle by the combination of 5 vertical lines and 5 horizontal lines, verify the aspect ratio and size, and determine the rectangle closest to the plate. If no candidate rectangle is established, no detection is made.

Fig.3 shows how plate candidate areas are extracted from the input image. The region of the vehicle is extracted as a candidate of the object region by Faster-RCNN [21] or the like. This is also demonstrated by Kido [22] and shows good results. Also, YOLO [23] can be used. The vehicle area is normalized to a size of  $400 \times 300$  pixels as shown in Fig.4. An area  $200 \times 75$  pixels including the plate candidate is set in this area. An example of the detected plate is shown in Fig.5.

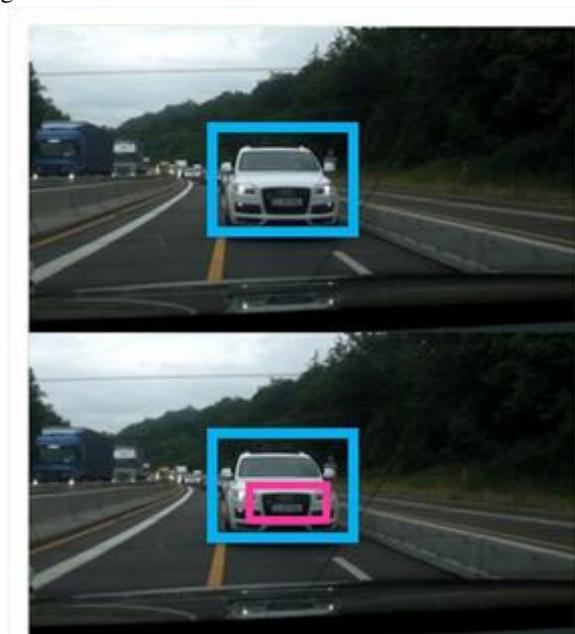


Fig.3 Vehicle area (upper) and license plate candidate area (lower).

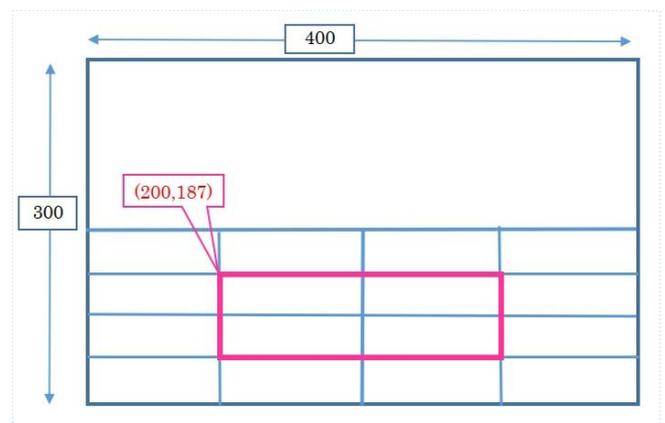


Fig. 4 Whole vehicle area(400x300) and candidate area of license plate(200x75).

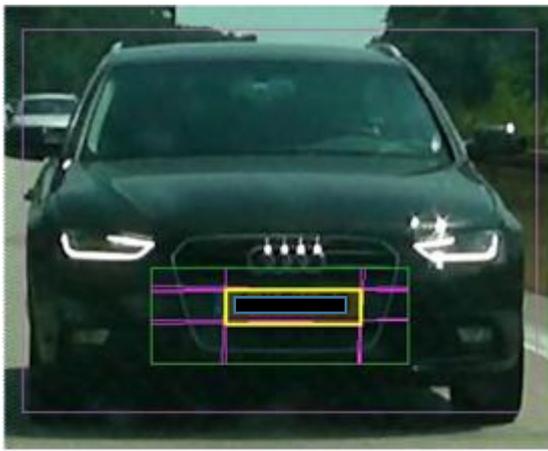


Fig.5 An example of detected plate in this paper.

IV. MACHINE LEARNING FOR THRESHOLD VALUE

Machine learning for threshold value applied to input image to get binary images is introduced. The binary images are important to be used in the following process. The threshold value was at the beginning of this project, the center value of luminance which is 127 or 128. Then, taking into account of input change of luminance, average value of input luminance or shifted average value by a fixed value are tried. But through experiments, detection ratio is affected by changing threshold values. The threshold may be affected by input environment which is beyond average luminance. To cope with this problem, automatic learning by machine learning tool is effective. In this paper a machine learning method is newly introduced to derive more effective threshold from input whole image, vehicle region image, or plate candidate region.

Fig.6 shows the upper and lower two areas of the vehicle area and the lower area divided into 16 areas. As the explanatory variables used in machine learning, the average value and the variance value of the luminance of the area numbered in Fig. 6 are used. The threshold value was changed as the correct value for supervised learning, and the value at which the software was able to detect the plate normally was allocated. Therefore, a predicted value of threshold is set as an objective function. If the correct data

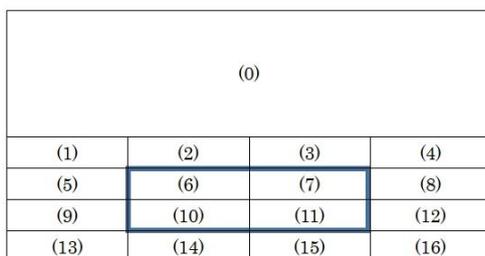


Fig.6. shows the upper(0) and lower(1)-(16) two areas of the vehicle area. The lower area divided into 16 areas.

has a width, the center value is taken as the correct value of the objective function. As a whole, machine learning system with at most 34 explanatory variables can be configured. Table 2 shows the sequence of data used. ML is a regression that predicts a threshold value, and in this case, a frequently used linear multiple regression function is used.

Table 2 Data structure for training and test.

image	ave0	ave1	ave2	.....	sd0	sd1	.....	sd16	thr
data_0	124.6	107.8	111.1	.....	.....	46	.....	68.8	85
data_1	88.54	112.6	54.65	.....	.....	73	.....	24.5	83
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
data_n	174.6	249.3	180.6	.....	.....	20	.....	59.2	147

V. EXPERIMENTS

The experiment was divided into three types.

- (1) Basic method: In the case of a fixed threshold as the basic configuration,
- (2) ML method: When ML is introduced and threshold is automatically determined,
- (3) Otsu method: The case where the famous Otsu's method is used as a method of binarization, will be described.

As the input image, a road image taken using an on-vehicle camera and a vehicle image on the road synthesized by software for performing a virtual test were used. The road images are selected from the real images [24] on the road in the front and rear of a vehicle traveling in Europe, especially in Germany (see Fig. 5). The synthesized images were generated using IPG Automotive's CarMaker® [25], which is widely used as Virtual Test Software (Fig. 6). Images are collected that are difficult to detect. Also, the size is Hi-definition (HD: 1920x1080) standard. The correct value "thr" (in Table 2) which is teacher data was obtained by changing threshold over the entire range for each image. In this preliminary experiment to obtain correct data, the median value of the threshold value when normal plate detection can be performed is determined, and up to 66 correct data can be obtained at present(n=66 in Table 2).

Table 3 is the result of performing a detection experiment by JavaANPR of the prior art [9]. In the case of HD size, in the first column, in the second column when vertical and horizontal are reduced to 1/2, and in the third column, the vehicle region size (400 × 300) is cut out. Because we use JavaANPR's software as is in Table 3, it is not possible to compare the same lines simply with the detection situation proposed in this paper. The images used in this paper are with increased levels of difficulty, such as those that are shot in backlight, dark images, bright images, artificial synthesized images by CarMaker®, and so on. Common

road images achieved precision = 0.923 in the old version of CT5L in the experiment by Okunuki.

Table 3 results by JavaANPR [9]

	1	2	3
images	original full_size 1920x1080	quarter 960x540	cropped 400x300
TP	0.303030303	0.015151515	0.272727273



Fig.6 A car Image generated by IPG CarMaker

(1) Basic method: CT5L method was executed with threshold fixed. In the experiments up to now, as the threshold value, for example, the intermediate value (127 or 128) of luminance, a value of 95 or the like have been used in the preliminary experiment with a fixed value. In this Basic experiment, changes in the detection rate of the entire search were examined in which the threshold was changed from 0 to 255 on the obtained input image. As a result, as shown in Fig.7, using 89 as the fixed threshold value was the best result. As for the other values, it can be seen that the detection rate randomly fluctuates and gradually decreases as it gets away from 89. Also, there is a second peak at 104, 105.



Fig.7 The case of fixed threshold

(2) ML method: The implementation of ML used the Linear Regression function of Scikit-learn included in the Library of Python. The explanatory variable is extracted

from the numerical data described in Section 4, and learning is performed by Linear\_Regression. When the predicted value of threshold output as a result is included in the range of the correct answer, it is regarded as the correct answer. That is to minimize the sum of squares of errors with the correct answer. The learning used Leave-One-Out Cross-Validation method (LOOCV) [26] which is a type of cross-validation [27]. In LOOCV, one test data is separated from learning data, and all remaining data are used for learning. Using the regression coefficient after learning, test data that has been isolated without being used for learning is input, a predicted value is determined, and the accuracy rate is measured. First of all, the performance was evaluated when the explanatory variable was limited to one. This is useful because it evaluates the effectiveness of each explanatory variable and is also a source of feature variable selection in subsequent multiple regression experiments.

Fig.8 shows the result when there is one explanatory variable. Using the luminance average value as an explanatory variable is better than using the standard deviation value, and the range is about 0.46 to 0.6. COD is also better for average than that for Sd. Also, as the position of the variable, 2, 6, 10, etc. are bad, and 4, 5, 8, 9, 11, 12, 13, etc. are good in results.

he combination which takes out several explanatory variables from 34 explanatory variables becomes huge. In fact, the total number of combinations that extract n from 34 is  $2^{34}-1$ , if n = 34 from the formula  $2^n = \sum_{k=0}^n nCk$ . Taking out multiple explanatory variables is called "Feature Selection", and methods to reduce the number of searches have been studied for a long time. The method of selecting the one with good performance in the case of one explanatory variable and sequentially increasing the combination is called "Forward Method" [28]. In addition, when increasing sequentially, it is considered as a bad effect that it can not be removed after registration once, and a flexible method of adding p pieces and excluding r pieces

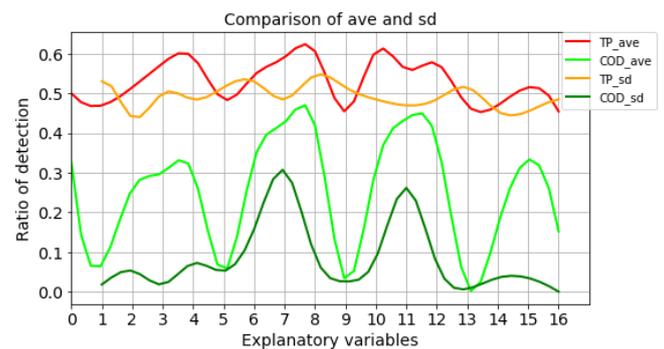


Fig.8 The case of a single explanatory variable.

For each explanatory variable, detection ratio is evaluated by average or variance. TP\_ave=True positive rate, ave means average of detection ratio, sd=Standard deviation, and COD=Coefficient Of Determination. Other than grid points are interpolated by the spline function. The same is true for the graphs below.

has been proposed as "+p-r method" [29]. Here, based on the Forward method, while increasing sequentially from the high precision explanatory variables, we will try to replace as appropriate when increasing the number. Similarly, Fig. 9., Fig. 10, Fig. 11, Table 4, Table 5 and Table 6 show the results when 3, 4, 5, 6 and 7 explanatory variables are used.

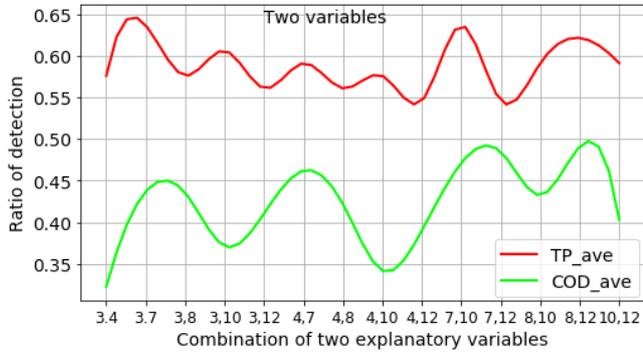


Fig.9 Results when using two explanatory variables.

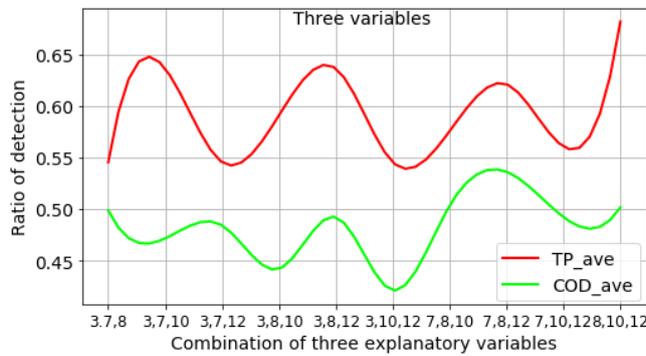


Fig.10 Results when using three explanatory variables.

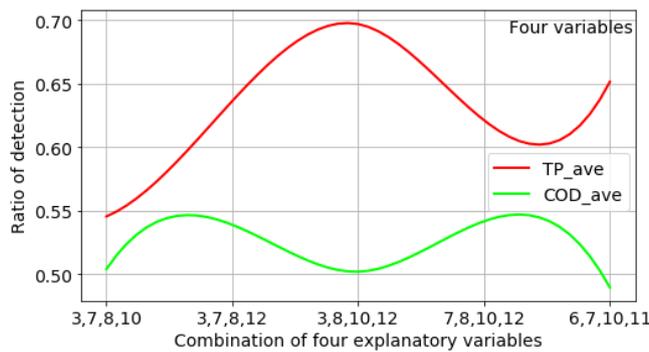


Fig.11 Results when using four explanatory variables.

Table 4 the case of Five Variables.

rate	ast6,ast7,ast10,ast11,ast12
TP	0.621212121
COD	0.514388478

Table 5 The case of six variables.

rate	ave4,ave6,ave7,ave10,ave11,sd11
TP	0.560606061
COD	0.543577619

Table6 Two cases of seven and eight variables.

The case of seven variables	
rate	ave5,ave6,ave7,ave8,ave9,ave11,sd11
TP	0.742424242
The case of eight variables	
rate	ave5,ave6,ave7,ave8,ave9,ave10,ave11,sd11
TP	0.712121212
COD	0.695194388

(3) Otsu method: There is a famous Otsu's method as a scheme of binarization. Here, image data is given, and experiments are performed in the case of plate detection using a threshold value obtained by the Otsu binarization method [29\_Otsu] as a predicted value. Otsu's method is to automatically find the optimum threshold iteratively so as to maximize the ratio of the interclass variance divided by the intraclass variance. The implementation of Otsu's binarization used the threshold function in OpenCV2. The image was given in three ways: an original image (1920x1080), a vehicle area (400x300) and a plate candidate. The fixed threshold used in the past (95, etc.) has been improved in some cases, but has not reached a good example by ML.

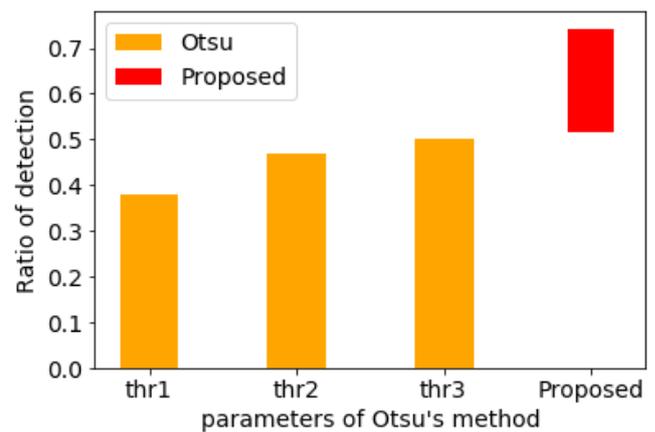


Fig. 12 ratio(TP) vs. area of threshold calculation. thr1=1920x1080, thr2=960x540, thr3=400x300.

## VI EVALUATION AND CONCLUSION

Table 7 shows the best results of for fixed threshold values case, for Otsu's automatic optimization method case and for proposed Machine Learning method case. In Otsu's

method, in addition to the automatically determined threshold value, the value when improved manually is added. In ML, the proposed method with seven variables at present is the best, and it is improved by 0.227 points in rate representation compared to the conventional fixed case.

In the field of deterministic verification, that is, in the field of the method that does not use deep learning at present, optimization of threshold is performed by fixed ML in the method of detecting features of the plate region. It is possible to construct the method which surpasses the conventional automatic optimization method Otsu Method and JavaANPR.

In the future, we will investigate the possibility of further performance improvement, such as mitigation for over-detection from the data distribution obtained during optimization. We will also incorporate and implement Virtual Test's LPD detection function as reference software.

Table 7 Comparison of the best value of three methods

	Fixed threshold	Otsu Method	ML(proposed)
param.	thr=89	thr3(200x75)	ave5,6,7,8,9,11, sd11
rate (TP)	0.515152	0.5 (automatic) 0.68(plus manual adjustment)	0.742424242

## REFERENCES

- [1] Documentation\_1st\_RoundTableAutonomousDriving.pdf <https://www.hs-kempten.de/fileadmin/fh-kempten/HK/news/2016/>
- [2] Kazuo Ohzeki, Yoshikazu Kido, Yutaka Hirakawa, Stefan Schneider, "Multi-Module Deep Learning Using Training Data from the first-Stage Error -- Effective For License Plate Detection --", IEICE Technical report vol. 117, no. 514, PRMU2017-176, pp. 25-30, March 2018, (in English)
- [3] Xiaowei Huang, Marta Kwiatkowska, Sen Wang and Min Wu, "Safety Verification of Deep Neural Networks", Keynote of CAV2017 July 2017 Heidelberg
- [4] Alexander, Robert David, Ashmore, Rob and Banks, "The State of Solutions for Autonomous Systems Safety.", White Rose Research Online, February 2018
- [5] James H. Gawron, Gregory A. Keoleian, Robert D. De Kleine, Timothy J. Wallington, and Hyung Chul Kim, "Life Cycle Assessment of Connected and Automated Vehicles: Sensing and Computing Subsystem and Vehicle Level Effects", Environ. Sci. Technol., 2018, 52 (5), pp 3249-3256, American Chemical Society Feb. 2018.
- [6] "Das Bundesverfassungsgericht" (in German). Bverfg.de. 3 November 2008. Retrieved 16 February 2009.
- [7] [https://en.wikipedia.org/wiki/Automatic\\_number\\_plate\\_recognition#Germany](https://en.wikipedia.org/wiki/Automatic_number_plate_recognition#Germany)
- [8] Cláudio Rosito Jung and Rodrigo Schramm, "Rectangle Detection based on a Windowed Hough Transform", Computer Graphics and Image Processing, 2004. Proceedings. 17th Brazilian Symposium, pp. 113 - 120, 17-20 Oct. 2004,
- [9] Ondrej Martinsky, "Algorithmic and mathematical principles of automatic number plate recognition systems", B.SC. THESIS, BRNO University of Technology, 2007.
- [10] Di Zang, Zhenliang Chai, Junqi Zhang, Dongdong Zhang, and Jiujun Cheng, "Vehicle license plate recognition using visual attention model and deep learning" Journal of Electronic Imaging SPIE 24(3) 033001-pp.1-10. May/June 2015.
- [11] S'ergio Montazzolli and Claudio Jung, "Real-Time Brazilian License Plate Detection and Recognition Using Deep Convolutional Neural Networks", 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), pp.55-62, Oct. 2017.
- [12] M. Dong, D. He, C. Luo, D. Liu, W. Zeng, "A CNN-Based Approach for Automatic License Plate Recognition in the Wild", 28th British Machine Vision Conference BMVC pp.1-12, Sept 2017.
- [13] S.G. Kim, H.G. Jeon and H.I. Koo, "Deep-learning- based license plate detection method using vehicle region extraction", Electronics Letters Vol. 53, Issue: 15, 7 20 Institution of Engineering and Technology, pp.1034-1036, 2017.
- [14] Hui Li, and Chunhua Shen, "Reading Car License Plates Using Deep Convolutional Neural Networks and LSTMs", arXiv:1601.05610 Jan, 2016.
- [15] Dehghan, Afshin; Zain Masood, Syed; Shu, Guang; Ortiz, Enrique G. (19 February 2017). "View Independent Vehicle Make, Model and Color Recognition Using Convolutional Neural Network". Archived from the original on 30 May 2018. Retrieved 30 May 2018 – via ResearchGate.
- [16] "OpenALPR Benchmarks". openalpr.com. 31 October 2017. <http://www.openalpr.com/benchmarks.html>
- [17] Laroca, Rayson; Severo, Evair; Zanlorensi, Luiz A.; Oliveira, Luiz S.; Resende Goncalves, Gabriel; Robson Schwartz, William; Menotti, David, "A Robust Real-Time Automatic License Plate Recognition Based on the YOLO Detector "International Joint Conference on Neural Networks (IJCNN), pp. 1-10. arXiv:1802.09567
- [18] Sérgio Montazzolli SilvaEmail, and Cláudio Rosito Jung, "License Plate Detection and Recognition in Unconstrained Scenarios", European Conference on Computer Vision, ECCV 2018: pp. 593-609, 2018
- [19] Kazuo Ohzeki, Takuya Okunuki, Stefan Schneider, "Number Plate Region detection using Luminance and Sobel filter data", The 18th Meeting on Image Recognition and Understanding (MIRU) DS1-9 pp.1-2 July 2015.
- [20] Kazuo Ohzeki, Takuya Okunuki, and Stefan Schneider, "AN ADVANCED NUMBER PLATE DETECTION METHOD FOR REAR-END COLLISION AVOIDANCE SYSTEM" Proc. Irish Transport Research Network (ITRN) Session 5b 27th-28th Aug. 2015.
- [21] Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", Arxiv:1506.01497 Part of: Advances in Neural Information Processing Systems 28 (NIPS 2015)
- [22] Yoshikazu kido, Yutaka Hirakawa, Kazuo Ohzeki, "Recognition of Driving Environment by Deep Learning Algorithm using adaptive Environment Model", Proc. PCSJ/IMPS P-5-15 Nov. 2017.
- [23] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi, "You Only Look Once: Unified, Real-Time Object Detection", IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Las Vegas June 2016 <https://arxiv.org/pdf/1506.02640.pdf>
- [24] <https://www.youtube.com/user/alg0z/videos>
- [25] <https://ipg-automotive.com/products-services/simulation-software/carmaker/>
- [26] Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, "An Introduction to Statistical Learning", Springer Texts in Statistics 2013.
- [27] Ron Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection", Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI-95 Contents Vol 2, Pages 1137-1143 Aug. 1995.
- [28] T. Marill and D. M. Green. "On the effectiveness of receptors in recognition systems", IEEE Transactions on Information Theory, 9:11-17, 1963.
- [29] P. Pudil, J. NovoviEova, J. Kittler, "Floating search methods in feature selection", Pattern Recognition Letters 15 (1994) 1 1 19-1 125.
- [30] Nobuyuki Otsu, "An Automatic Threshold Selection Method Based on Discriminant and Least Squares Criteria", ieice transaction D Vol.J63-D no.4 pp.349-356 April. 1980.

# Depth Map Improvements for Stereo-based Depth Cameras on Drones

Daniel Pohl  
Intel Corporation,  
Konrad-Zuse-Bogen 4,  
Krailling, Germany  
daniel.pohl@intel.com

Sergey Dorodnicov  
Intel Corporation,  
Rachel 4,  
Haifa, Israel  
sergey.dorodnicov@intel.com

Markus Achtelik  
Intel Corporation,  
Konrad-Zuse-Bogen 4,  
Krailling, Germany  
markus.achtelik@intel.com

**Abstract**—Using stereo-based depth cameras outdoors on drones can lead to challenging situations for stereo algorithms calculating a depth map. A false depth value indicating an object close to the drone can confuse obstacle avoidance algorithms and lead to erratic behavior during the drone flight. We analyze the encountered issues from real-world tests together with practical solutions including a post-processing method to modify depth maps against outliers with wrong depth values.

**Index Terms**—depth camera, stereo, computer vision

## I. INTRODUCTION

In the last decade, depth cameras have become available in more affordable versions which increased the usage both in the industrial as well as in the consumer space. One interesting use case is on drones, where stereo-based depth cameras generate data for obstacle avoidance algorithms to keep the drone safe. However, the algorithms used to calculate depth information are trying to solve an under-determined problem. From two-dimensional images, data in three dimensions is reconstructed. Therefore, it seems only natural that in certain cases the generated depth images might contain wrong data as shown in Figure 1. Specifically, when used in larger outdoor environments at different weather conditions like drones would exhibit, the set of parameters and requirements might be very different from other common use cases for depth sensors as found in indoor scenarios like finger tracking or gesture recognition.

In this paper, we present the encountered challenges of using depth cameras on drones and how we overcame them. **Our contributions are:**

- Description of stereo camera issues on drones
- Solutions to minimize the encountered problems
- Release of the solutions as highly optimized open source code

In the following, we will first give an overview of related work in the space of drones with different depth cameras. Next, we describe the use case of the depth sensor on our drone and issues observed for enabling automatic obstacle avoidance. After specifying the hardware and software system, we take a look at incorrect depth values from the used depth cameras. Having all of this laid out, we provide improvements against depth outliers through a variety of methods like calibration, depth camera settings and post-processing methods.



Figure 1. The top image shows the depth map from the scene at the bottom. The colors are applied depending on the distance in meters as shown in the scale on the right part of the image. At the light gray wall with thin horizontal stripes, the algorithm of the depth sensor wrongly estimates an object close to the camera.

We compare the results of the post-processing steps and provide a performance analysis of the used algorithms. We discuss current limitations and give an outlook on further improvements. Last, we conclude and link to our open source implementation.

## II. RELATED WORK

There are various devices to measure depth to other objects. Options which have also been used on drones include ultrasonic [1], lidar [2], [3], radar [4] or depth camera-based systems [5]–[8]. While all of these have their advantages and drawbacks, we focus in this work on depth camera-based

systems due to their light weight, detailed depth information and relatively low cost.

In the category of depth cameras, we describe two very common types and their differences [9].

**Time Of Flight (TOF) cameras:** a laser or LED is used to illuminate where the camera is pointing at [10]. As the constant speed of light is known, the round-trip time of such a light signal returning to the camera sensor can be used to calculate an approximate distance. Common advantages of these depth sensors are simplicity, efficient distance algorithm and their speed. Their drawbacks show in bright outdoor usage where the background light might interfere with measurements, potential interference with other TOF devices and issues at reflections.

**Stereo-based depth sensors:** these devices are taking two images with a fixed, known offset between the two image cameras. Using stereo matching algorithms [11], [12] together with the known intrinsic and extrinsic parameters of the camera, they can generate approximate depth values for the image. Usually, these sensors consume less power compared to TOF cameras.

### III. DRONE USE CASE

To avoid accidents, injuries and crashes, it is very important for drones to avoid flying into obstacles. Depth cameras help the drone to "see" the environment. The obstacle avoidance algorithms that we use are taking the depth image from one or more depth cameras. As we know the mounted camera position and orientation on the drone and the GPS location of the drone, we transform the data from the depth image into world space. We map those depth values to 3D voxel locations. For the voxel value, we update the probability of that space to be occupied. Having the voxel map available, we check the drone's heading and velocity against potential obstacles in that direction. If we find any, we redirect to avoid a collision.

As we found in real-world usage of drones with depth sensors, there are sometimes issues that the depth values are not correct and can therefore lead to problems. For example, suddenly, a wrong, very close depth value appears in front of the drone. This might be interpreted as an obstacle to which our safety distance is not kept and strongly violated. A common reaction might be to move the drone quickly away from that obstacle or to at least not move further into the direction of the obstacle. For a drone operator on the ground observing what happens in the sky, such behavior of the drone is not comprehensive. The operator sees that there is no obstacle, yet the drone behaves in an undesired way trying to avoid invisible objects.

### IV. SYSTEM

In the following scenarios, we use the Intel NUC7i7BNH platform with the Intel Core i7-7567U (2 cores, 4 threads) at a base frequency of 3.5 GHz with 16 GB memory. Given the requirements of being able to work outside in bright environments and the goal of having a low power consumption, we decided to use a stereo-based depth sensor. The model is Intel RealSense [13] D435i with the firmware 5.11.6.250.



Figure 2. Drone with depth sensor

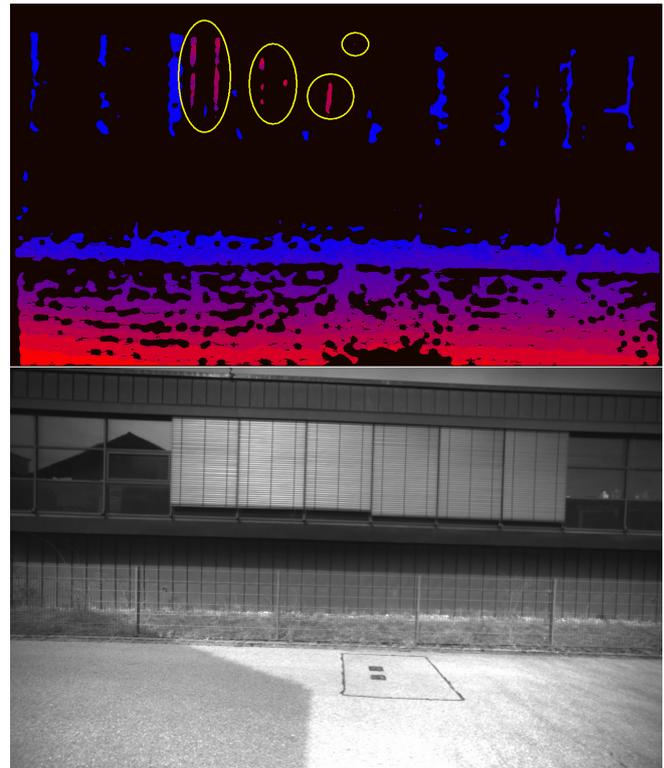


Figure 3. A case in which parts of the blinds on the windows are wrongly indicating depth which is very close to the camera. Near distances are represented in an intense red, while the farther away it gets, the coloring changes to blue.

The system runs Ubuntu 18.04 with the Intel RealSense SDK 2.0 (build 2.23.0). Some of the visualizations are generated with the RealSense Viewer 2.23.0. For image operations, we use OpenCV 3.4.5. The depth camera is mounted on an Intel Falcon 8+ octocopter (Figure 2). We use a camera resolution of  $848 \times 480$  pixels at 30 frames per second.

### V. INCORRECT DEPTH VALUES

As mentioned in Section III, we discovered some cases in which depth values in the depth map were not accurate and disturb the obstacle avoidance algorithms. Figure 1 shows one example. We provide another case in Figure 3.

Both cases have in common that there is a structure with repetitive content which can easily disturb stereo feature matching as almost the same color values are frequently repeated in neighboring areas.

## VI. IMPROVEMENTS

In this section, we provide improvements for the previously described depth maps with some incorrect depth values.

### A. Calibration

Depth cameras are shipped with a previously executed factory calibration. Due to the stress on the modules endured by a potential air freight delivery with different pressure conditions at such high altitudes and potential shaking during transportation, it can happen that physical properties of the device slightly differ from the state it was during calibration.

At least in one case we found significant improvements when running a local calibration on the device. As test setup, we used a carpet intended for children to play with small toy vehicles on it. The carpet provides strong features which can be picked up by the stereo algorithm. For this test, we used very strict camera settings which rejected depth values if their confidence was not extremely high.

For the Intel RealSense D435i camera, there are tools that allow a recalibration within a few minutes. As shown in Figure 4, this can increase the confidence in depth values and therefore provide more valid inputs. In most real-world cases the differences will not be as high as illustrated here, but this shows how important an accurate camera calibration is.

### B. Depth Camera Settings

As a guideline for outdoor depth sensing as used on drones, we prefer having fewer depth values at a high confidence compared to receiving many values which are less certain to be valid. In the RealSense D435i camera, there are various settings affecting this which can be modified through visual tools like the RealSense Viewer and can be stored in .json files. Those configuration files can be uploaded in the application via API calls. We describe the most relevant changes in the settings that we made compared to the default. To give a better understanding of the parameters on the resulting images, we show different settings in the Appendix.

**texturecountthresh, texturedifferencethresh:** These settings describe how much difference in intensity in the gray scale stereo image needs to be to determine a valid feature. In outdoor usage, the sky and clouds provide an almost similar color with only small deviations. Walls captured during inspection flights might have areas of the same color which do not make strong features. To increase the confidence on depth values, we increased the values of *texturecountthresh*, which sets how many pixels of evidence of texture are required from 0 to 4 and set the value of *texturedifferencethresh*, how big a difference is required for evidence of texture, to 50.

**secondpeakdelta:** When analyzing the disparities of an area in the stereo images for a match, there might be one clear candidate indicating a large peak in terms of correlation. In some cases, multiple candidates could be viable at different peak levels. The second peak threshold determines how big the difference from another peak needs to be, in order to have confidence in the current peak being the correct one. We increased this value from a default of 645 to 775.

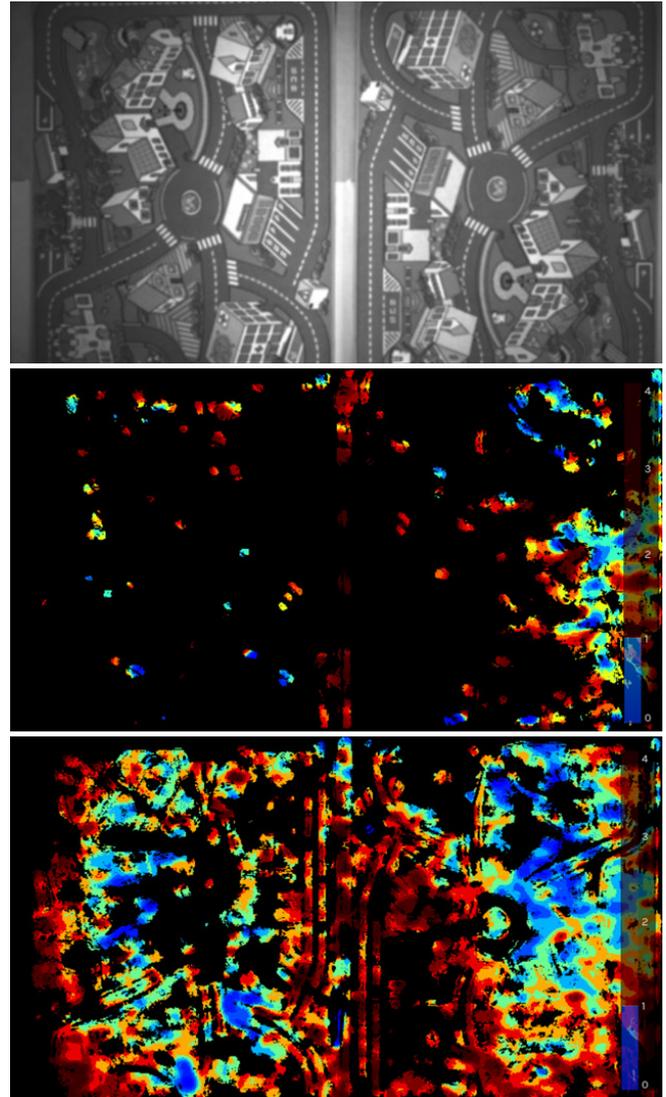


Figure 4. The top gray scale image shows the carpet for toys as target in a test setup. The middle image shows the depth values before manual calibration with camera settings for very high confidence of depth values. The bottom image shows the depth values after recalibration.

**scanlinep1, scanlinep1onediscon, ...:** For finding the best correlation form the disparities, a penalty model is used as described in [14]. In addition to estimating the validity of a current correlation, neighboring areas with their estimate are analyzed and taken into account. A small difference can be expressed in a small penalty (*scanlinep1* = 30) while a larger difference leads to a second penalty value (*scanlinep2* = 98). Both penalties are added together in an internal cost model for the likelihood of a correlation to be the correct one. Further fine tuning on large color or intensity differences between the left and right image can be set with *scanlinep1onediscon* and *scanlinep1twodiscon*.

**medianthreshold:** When looking for a peak regarding correlation, we want it to have a significantly large value

to clearly differ from the median of other correlation values. While the default is set to 796, we found that we were able to lower this value safely to 625. This did not introduce any noticeable artefacts, but made more valid depth values available.

**autoexposure-setpoint:** The autoexposure setting can be changed to deliver a darker (lower value) or brighter image. It is set to 1500 by default. For outdoor usages, we found the brighter value of 2000 to work better. Details in the sky like clouds are not relevant for us, so if this part is overexposed, it has no negative effect. On the positive side, increasing brightness makes darker objects like the bark on a tree brighter and enables better feature detection on it.

We present the full .json file with all settings in the Appendix.

### C. Post-processing of Depth Images

With a good depth camera calibration and the modified parameters, we are able to get good images with relatively high confidence features. However, for our purpose this is still not enough and cases with invalid depth values have still been observed. We tried many different other parameter settings, but in the end, we were not able to remove the outliers just through parameters without losing almost all other valid depth data. Instead, to handle the invalid depth values, we are applying post-processing steps to the received depth image. As described in [15], there are various known methods for post-processing like downsizing the image in certain ways to smooth out camera noise, applying edge-preserving filtering techniques or doing temporal filtering across multiple frames.

In our outdoor drone use case, we apply different post-processing methods. For the ones we describe, we additionally require reading out the left rectified camera image stream which is synchronized with the depth image. In our depth camera model this image is in an 8-bit gray scale format. The pseudo-code for our post-processing operations is in this listing:

```

1  const int reduceX = 4;
2  const int reduceY = 4;
3  cvResizeGrayscaleImage(reduceX, reduceY);
4  resizeDepthImageToMinimumInBlock(reduceX, reduceY);
5
6  // create edge mask
7  cvScharrX(grayImage, maskEdgeX);
8  cvScharrY(grayImage, maskEdgeY);
9
10 convertScaleAbsX(maskEdgeX);
11 convertScaleAbsY(maskEdgeY);
12
13 cvAddWeighted(maskEdgeX, maskEdgeY, maskEdge, 0.5);
14 cvThreshold(maskEdge, 192, 255, THRESH_BINARY);
15
16 // create corner mask
17 cvHarris(grayImageFloat, maskCorners, 2, 3, 0.04);
18 cvThreshold(maskCorners, 300, 255, THRESH_BINARY);
19
20 // combine both masks
21 cvBitwiseOr(maskCombined, maskEdge, maskCorners);
22
23 // apply morphological opening
24 cvMorphOpen(maskCombined, MORPH_ELLIPSE(3, 3));
25
26 // use mask on depth image
27 depthImage.cvCopy(depthImageFinal, maskCombined);

```

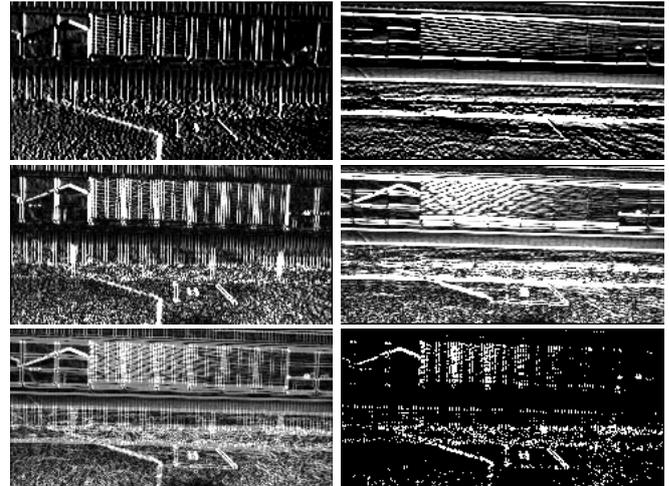


Figure 5. Steps for creating the edge mask. The top row shows the Scharr images in X and Y dimension. The second row applies the absolute function to the values from the top row. The last row shows left the added images from the middle row. On the right, it shows the final mask with the applied binary threshold function.

In the lines 1-4, we are downsizing both the depth and the camera image by a factor of four in each dimension. For the depth map, we search within a 4x4 pixel block for the closest depth value which is not zero, meaning not invalid. We take this value as the downsized pixel value. The reason for this selection is that for obstacle avoidance our most important information is which object might be the closest to us. For the gray scale image, we can use regular OpenCV downsizing. In our case, nearest-neighbor downsizing was sufficient, but, depending on the performance budget, bilinear filtering might be chosen as well. After resizing, the depth image and camera image have been lowered from a resolution of  $848 \times 480$  pixels to  $242 \times 120$  pixels. With 16 times fewer pixels, further processing on the images will be much faster.

Edges and corners are very robust features for stereo matching. To achieve even higher confidence in the depth values, we want to mask out all depth values which do not have edges or corners in the corresponding area of the gray scale image. To do this, we create an image mask for edges and one for corners. For edge detection, we use the OpenCV Scharr operator [16] as shown in lines 7 and 8 of the pseudo-code listing. For the intermediate images in X and Y dimension, we apply the absolute function and convert them into an 8-bit format (line 10, 11). We add both images together and apply a binary threshold on the mask (lines 13, 14). Using the case from Figure 3, we visualize these processing steps in Figure 5.

For creating the corner mask, we use the Harris Corner Detector [17] in OpenCV (line 17). Again, we apply a threshold in the line below. We combine the mask for edges with the mask for corners in line 21. To eliminate too small areas in the mask, we apply the morphological opening operation on the mask which applies an erosion followed by a dilation on the image (line 24). We apply the final mask to the resized depth image. Only where positive values are in the mask, the

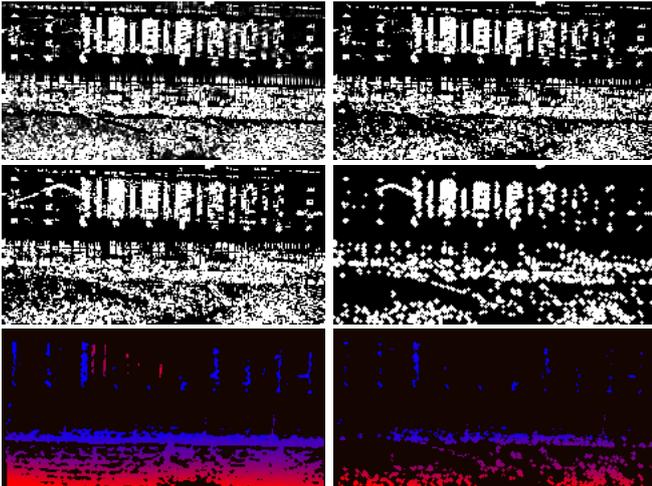


Figure 6. Steps for creating the corner mask. Top left shows the corners as detected by Harris. On the right, the binary threshold is applied to that. In the second row on the left, the combined edge and corner mask is shown. On the right side the final mask is shown after the opening function has been applied. The bottom row shows left the original, resized depth map. On the right, the mask has been applied to it. This is the final version of the depth map without outliers of wrong depth.

depth value will be copied into the final image, otherwise it will be set to zero, indicating no valid depth information (line 27). We visualize these steps in Figure 6.

#### D. Results

Using recalibration, tuning of the depth camera parameters and applying the post-processing steps as described, we got a much higher quality depth map as the example with the result in Figure 6 (bottom right) shows. A comparison between the before and after images with a resolution of  $242 \times 120 = 29040$  pixels, is shown in Table I.

Table I

COMPARING THE BEFORE AND AFTER IMAGE FROM FIGURE 6 (BOTTOM). THE FIRST NUMBER INDICATES THE AMOUNT OF PIXELS IN AN IMAGE WITH A RESOLUTION OF  $242 \times 120$  PIXELS. THE SECOND NUMBER SHOWS THE PERCENTAGE OF PIXELS IN THAT IMAGE.

	original	our method
number of depth values	7375 (25%)	2802 (10%)
number of outliers	94 (0.3%)	0 (0%)

While the image loses more than half of its valid depth information with our method, it also eliminates all outliers. As it can be seen in comparing both images, the loss happens relatively evenly across areas. For our obstacle avoidance this means that we still have enough information in these areas to be aware of potential objects in our path. To repeat the statement we made before: we prefer having fewer depth values at a high confidence compared to receiving many values which are less certain to be valid. Using our method, this goal is achieved.

We tested our method on multiple hours of log files from various drone flights. In almost all cases, we were able to

filter out wrong depth measures that would have impacted the drone’s obstacle avoidance to work correctly.

#### E. Performance

The post processing steps will increase the required compute load. We optimized our code to make use of AVX2 functions for our custom-written resizing function which we make available as open source. OpenCV, compiled with the right flags, will use AVX2 intrinsics for the relevant functions. We measured how much time the individual steps for post-processing took for processing 30 frames (the amount of frames we receive within one second from the depth camera) and show this in Table II.

Table II  
TIME IN MS FOR POST-PROCESSING STEPS FOR 30 FRAMES.

resize gray scale	1.2
resize depth map	1.5
create edge mask	2.9
create corner mask	9.3
combine masks	1.8
opening mask	1.3
apply mask	0.3

**total** 18.3

## VII. LIMITATIONS AND OUTLOOK

There are still some rare cases in which wrong depth makes it through all the suggested methods. The area of pixels with wrong depth is already much smaller with our methods. To increase the robustness against these rare outliers, we recommend using the depth data in a spatial mapping like in a 3D voxel map. Popular libraries like Octomap [18] are a good starting point. Before values are entered into such a spatial structure, it might be required to have multiple positive hits for occupancy over multiple frames and/or observations of obstacles from slightly different perspectives. In the case of drones, movement is pretty common and even when holding the position, minimal movements from wind might already change what the depth camera delivers. The position of a wrong depth value and its corresponding 3D space might change by such a small movement. As the incorrect depth values are not geometrically consistent, they might be filtered out through the spatial mapping technique.

While the performance impact of our routines is already relatively low for a modern PC-based system, the overhead might still hurt performance on highly embedded systems. In future versions of depth cameras, it might be a desired step to have our described methods directly implemented in hardware.

## VIII. CONCLUSION

In this work, we described the issues of receiving wrong depth data that was observed in some drone flights outdoors. Through proper calibration, modification of internal depth camera parameters and a series of post-processing steps on the depth map, we were able to clean up almost all outliers with wrong

depth. The resulting depth data can be used for robust obstacle avoidance with spatial mapping of the environment. Our highly optimized algorithms for post-processing are released as open source under <https://github.com/IntelRealSense/librealsense>.

#### REFERENCES

- [1] N. Gageik, T. Müller, and S. Montenegro, “Obstacle Detection and Collision Avoidance using Ultrasonic Distance Sensors for an Autonomous Quadcopter”, *University of Wurzburg, Aerospace information Technology Wurzburg*, pp. 3–23, 2012.
- [2] L. Wallace, A. Lucieer, C. Watson, and D. Turner, “Development of a UAV-LiDAR System with Application to Forest Inventory”, *Remote Sensing*, vol. 4, no. 6, pp. 1519–1543, 2012. DOI: 10.3390/rs4061519.
- [3] A. Ferrick, J. Fish, E. Venator, and G. S. Lee, “UAV Obstacle Avoidance using Image Processing Techniques”, in *IEEE International Conference on Technologies for Practical Robot Applications (TePRA)*, 2012, pp. 73–78. DOI: 10.1109/TePRA.2012.6215657.
- [4] K. B. Ariyur, P. Lommel, and D. F. Enns, “Reactive Inflight Obstacle Avoidance via Radar Feedback”, in *Proceedings of the 2005 American Control Conference*, IEEE, pp. 2978–2982. DOI: 10.1109/ACC.2005.1470427.
- [5] K Boudjit, C Larbes, and M Alouache, “Control of Flight Operation of a Quad rotor AR. Drone Using Depth Map from Microsoft Kinect Sensor”, *International Journal of Engineering and Innovative Technology (IJEIT)*, vol. 3, pp. 15–19, 2013.
- [6] A Deris, I Trigonis, A Aravanis, and E. Stathopoulou, “Depth cameras on UAVs: A first approach”, *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42, p. 231, 2017. DOI: 10.5194/isprs-archives-XLII-2-W3-231-2017.
- [7] I. Sa, M. Kamel, M. Burri, M. Bloesch, R. Khanna, M. Popovic, J. Nieto, and R. Siegwart, “Build Your Own Visual-Inertial Drone: A Cost-Effective and Open-Source Autonomous Drone”, *IEEE Robotics & Automation Magazine*, vol. 25, no. 1, pp. 89–103, 2018. DOI: 10.1109/MRA.2017.2771326.
- [8] S. Kawabata, K. Nohara, J. H. Lee, H. Suzuki, T. Takiguchi, O. S. Park, and S. Okamoto, “Autonomous Flight Drone with Depth Camera for Inspection Task of Infra Structure”, in *Proceedings of the International MultiConference of Engineers and Computer Scientists*, vol. 2, 2018.
- [9] H. Sarbolandi, D. Lefloch, and A. Kolb, “Kinect Range Sensing: Structured-Light versus Time-of-Flight Kinect”, *Computer vision and image understanding*, vol. 139, pp. 1–20, 2015. DOI: 10.1016/j.cviu.2015.05.006.
- [10] P. Zanuttigh, G. Marin, C. Dal Mutto, F. Dominio, L. Minto, and G. M. Cortelazzo, “Time-of-Flight and Structured Light Depth Cameras”, *Technology and Applications*, 2016. DOI: 10.1007/978-3-319-30973-6.
- [11] S. T. Barnard and M. A. Fischler, “Computational Stereo”, 1982. DOI: 10.1145/356893.356896.
- [12] T. Kanade and M. Okutomi, “A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment”, in *Proceedings. 1991 IEEE International Conference on Robotics and Automation*, pp. 1088–1095. DOI: 10.1109/ROBOT.1991.131738.
- [13] L. Keselman, J. Iselin Woodfill, A. Grunnet-Jepsen, and A. Bhowmik, “Intel RealSense Stereoscopic Depth Cameras”, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 1–10. DOI: 10.1109/CVPRW.2017.167.
- [14] M. Michael, J. Salmen, J. Stallkamp, and M. Schlipsing, “Real-time Stereo Vision: Optimizing Semi-Global Matching”, in *IEEE Intelligent Vehicles Symposium*, 2013, pp. 1197–1202. DOI: 10.1109/IVS.2013.6629629.
- [15] A. Grunnet-Jepsen and D. Tong, *Depth Post-Processing for Intel RealSense D400 Depth Cameras*, <https://www.intel.com/content/dam/support/us/en/documents/emerging-technologies/intel-realsense-technology/Intel-RealSense-Depth-PostProcess.pdf>.
- [16] H. Scharr, “Optimale Operatoren in der digitalen Bildverarbeitung”, 2000. DOI: 10.11588/heidok.00000962.
- [17] K. G. Derpanis, “The Harris Corner Detector”, *York University*, 2004.
- [18] A. Hornung, K. M. Wurm, M. Bennewitz, C. Stachniss, and W. Burgard, “Octomap: An Efficient Probabilistic 3D Mapping Framework Based on Octrees”, *Autonomous robots*, vol. 34, no. 3, pp. 189–206, 2013. DOI: 10.1007/s10514-012-9321-0.

#### APPENDIX

To give a better overview of the impact of changing some of the mentioned RealSense depth camera parameters, we provide examples of the resulting images from Figure 7 to Figure 11. In order to find the best matching values, this was tested and fine-tuned on various environments: natural, industrial, residential and mixtures of those. The height was varied between looking at objects almost at the same height and from a much higher perspective, e.g. 30 to 50 meters above ground. When testing different parameters on the ground, we recommend using the Intel RealSense Viewer in which the parameters can be changed in real-time through sliders to directly see the impact on the images.



Figure 7. Gray scale images with different auto exposure values: 1500, 2000 (ours), 2500.

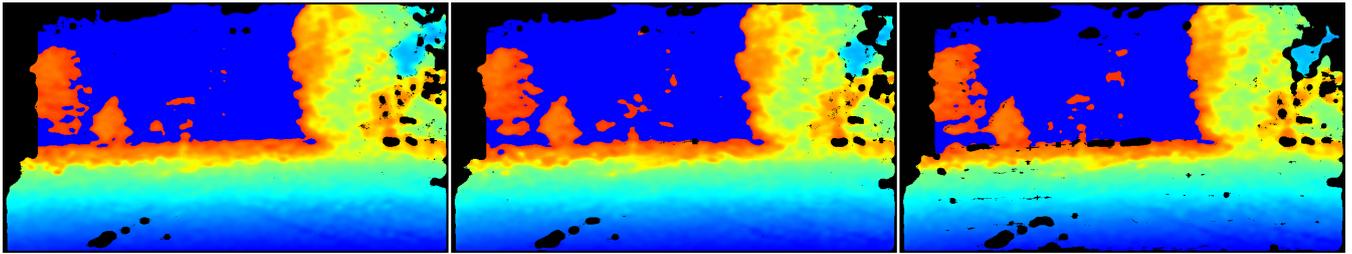


Figure 8. Depth images with different secondpeakdelta values: 400, 645, 775 (ours).

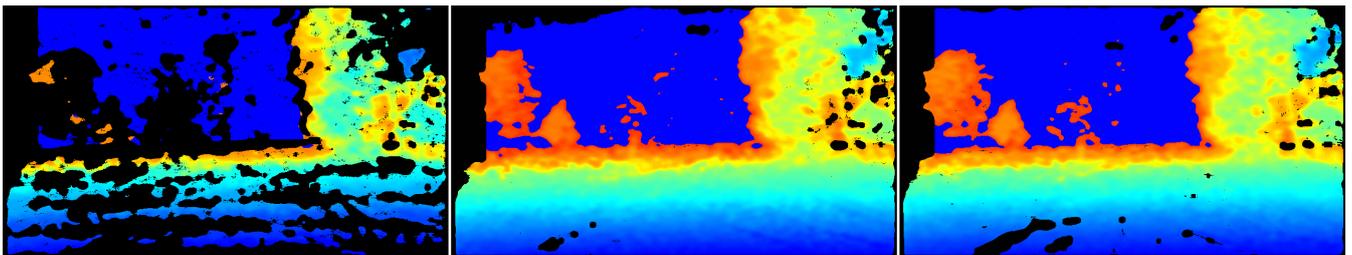


Figure 9. Depth images with different penalty values (scanlinep2onediscon): 50, 105 (ours), 235.

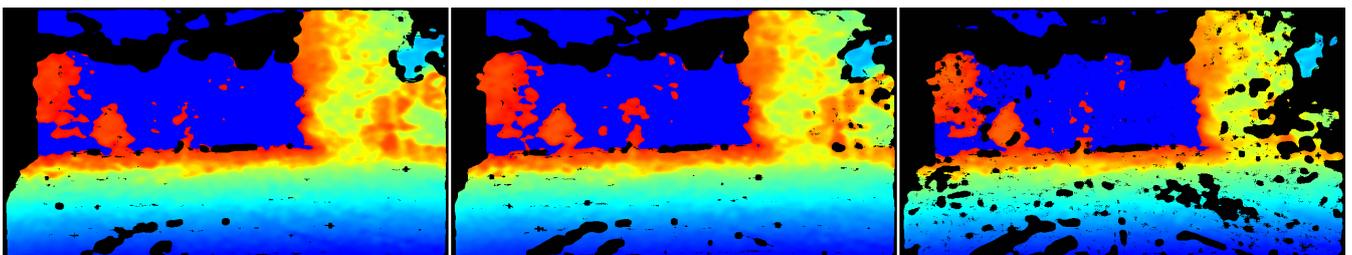


Figure 10. Depth images with different values for texturecountthresh and texturedifferencethresh: (0, 0), (4, 50) (ours), (8, 100).

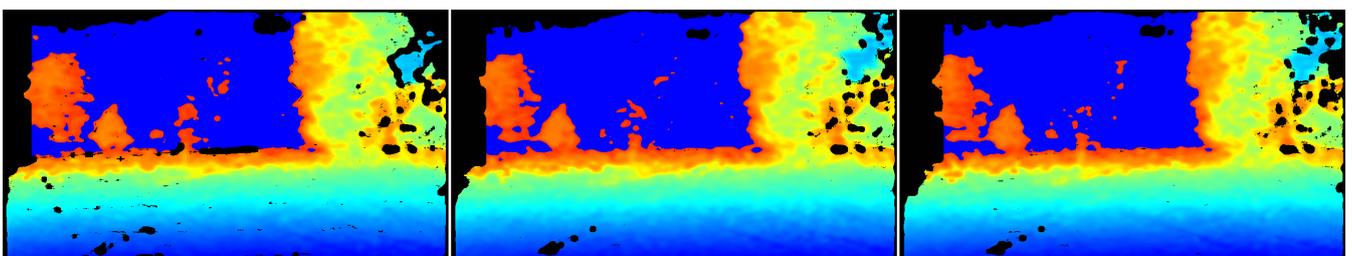


Figure 11. Depth images with different values for medianthreshold: 500, 625 (ours), 796.

The following is the text for the .json file that can be loaded in Intel RealSense tools and API calls to configure the cameras as described in the paper.

```

"aux-param-autoexposure-setpoint": "2000",
"aux-param-colorcorrection1": "0.298828",
"aux-param-colorcorrection10": "0",
"aux-param-colorcorrection11": "0",
"aux-param-colorcorrection12": "0",
"aux-param-colorcorrection2": "0.293945",
"aux-param-colorcorrection3": "0.293945",
"aux-param-colorcorrection4": "0.114258",
"aux-param-colorcorrection5": "0",
"aux-param-colorcorrection6": "0",
"aux-param-colorcorrection7": "0",
"aux-param-colorcorrection8": "0",
"aux-param-colorcorrection9": "0",
"aux-param-depthclampmax": "65536",
"aux-param-depthclampmin": "0",
"aux-param-disparityshift": "0",
"controls-autoexposure-auto": "True",
"controls-autoexposure-manual": "8500",
"controls-depth-gain": "16",
"controls-laserpower": "0",
"controls-laserstate": "on",
"ignoreSAD": "0",
"param-autoexposure-setpoint": "2000",
"param-censusenablereg-udiameter": "9",
"param-censusenablereg-vdiameter": "9",
"param-censususize": "9",
"param-censusvsize": "9",
"param-depthclampmax": "65536",
"param-depthclampmin": "0",
"param-depthunits": "1000",
"param-disableraucolor": "0",
"param-disablesadcolor": "0",
"param-disablesadnormalize": "0",
"param-disableslleftcolor": "0",
"param-disableslrightcolor": "1",
"param-disparitymode": "0",
"param-disparityshift": "0",
"param-lambdaad": "751",
"param-lambdacensus": "6",
"param-lefttrighthreshold": "10",
"param-maxscorethreshb": "1423",
"param-medianthreshold": "625",
"param-minscorethresha": "4",
"param-neighborthresh": "108",
"param-raumine": "6",
"param-rauminn": "3",
"param-rauminssum": "7",
"param-raumins": "2",
"param-rauminw": "2",
"param-rauminwesum": "12",
"param-regioncolorthresholdb": "0.784736",
"param-regioncolorthresholdg": "0.565558",
"param-regioncolorthresholdr": "0.985323",
"param-regionshrinku": "3",
"param-regionshrinkv": "0",
"param-robbinsmonrodecrement": "5",
"param-robbinsmonroincrement": "5",
"param-rsmdiffthreshold": "1.65625",
"param-rsmrauslodiffthreshold": "0.71875",
"param-rsmremovethreshold": "0.809524",
"param-scanlineedgetaub": "13",
"param-scanlineedgetaug": "15",
"param-scanlineedgetaur": "30",
"param-scanlinep1": "30",
"param-scanlinep1onediscon": "76",
"param-scanlinep1twodiscon": "86",
"param-scanlinep2": "98",
"param-scanlinep2onediscon": "105",
"param-scanlinep2twodiscon": "33",
"param-secondpeakdelta": "775",
"param-texturecountthresh": "4",
"param-texturedifferencethresh": "50",
"param-usersm": "1",
"param-zunits": "1000"

```

# Comparison of singing voice quality from the beginning of the phonation and in the stable phase in the case of choral voices

Edward Pótrolniczak

West Pomeranian University of Technology  
Faculty of Computer Science and Information Technology  
ul. Żołnierska 52, 71-210 Szczecin, Poland  
Email: epolrolniczak@wi.zut.edu.pl

Michał Kramarczyk

West Pomeranian University of Technology  
Faculty of Computer Science and Information Technology  
ul. Żołnierska 52, 71-210 Szczecin, Poland  
Email: mkramarczyk@wi.zut.edu.pl

**Abstract**—In the process of acoustic voice analysis, in this case of singing, it is important that the sound samples contain a stable phase of phonation. Sometimes, however, it is not possible. This study was prepared to determine how big are the differences between the values of the acoustic parameters obtained for the initial phase of phonation and for the stable phase of phonation. The values of acoustic parameters, such as, among others shimmer, jitter, RAP, PPQ, APQ, HNR or SPR were estimated for registered singing samples in the initial phase of phonation and in the middle phase. The analysis were performed over the samples of singing of the vowel 'a' recorded many times for different pitches. In the process of analyzing of the obtained results, it was found that the impact of the selection phase of phonation for analysis is crucial in assessing the singing voice quality.

## I. INTRODUCTION

**T**HE motivation for taking up the research on the singing voice acoustic parameters analysis was the need of assessment of singing quality. It may be useful for training lessons of voice production. It can be useful to help singers make a progress and it may allow for self-correction of selected voice parameters. It can be also very important for the choirs constantly working on the voice.

To analyse singing voice a must is to determine intonation or vibrato [1]. Some authors try to analyse the singing voice based on mel-cepstral features [2] or voice and speech features like Singing Power Ratio [3]. Anyway, there are many available acoustic parameters which may be investigated in the singing voice quality assessment.

One of the problems in the case of singing voice analysis is to obtain stable values of the parameters from the samples of singing. Due to the character of the singing signal envelope, it seems that determining the values based on the initial fragments of the singing recording may have an impact on the analytical process. So if it turns out that the values of the parameters from the beginning and the middle of the phrase differ significantly, it means that short signals for which a significant part is the attack and decay phase should not be applied to analysis.

During the creation of the database, the authors of this work paid attention to the quality of the samples. The recordings were carried out in appropriate conditions, the samples were subjected to precise segmentation. The samples obtained by us last for 3-4 seconds so they ensure that the middle part is the most valuable. In order to determine the analysed parameters, the samples were cut at an additional 5% from the beginning and the end. Regardless of these treatments, the authors were not sure if the samples throughout the entire run have similar quality or maybe the initial fragments are out of quality from the middle ones. This doubt was behind the undertaking of the described research and observations.

The voice, in general, is produced by a vocal instrument consisting of three elements: a breathing apparatus, oscillating vocal folds and a vocal tract. Breathing has a decisive impact on all activities related to voice emission. The entire phonation process can be represented using the ADSR (Attack-Decay-Sustain-Release) model that describes production of a single sound. It can be used for sound analysis [4], [5] and synthesis [6]. It is also the description of the sound waveform in the MIDI standard [7].

The sound attack is first part of the ADSR envelope. The attack ('on the sound') is used to modify the first phase of the amplitude envelope [8] of the generated sound, in which the sound gains the highest amplitude. This is followed by decay section during which the amplitude is reduced. As the next a sustain stage characterized by a stable pitch amplitude is visible. The ADSR envelope is completed at the release stage. The ADSR sections are illustrated in fig. 1.

The quality of singing is related to breathing. The proper breath before the phonation results in a good beginning of each phrase [9]. This is especially important in singing phrases that start with a vowel. The attack is more precise in the case of professional singers. Choral singers are less precise at this stage of voice production. The practice can solve this problem, but choir members usually develop their voice in groups, making their vocal abilities are similar in the group. It should be kept in mind that the presented investigation concerns the choral voices analysis.

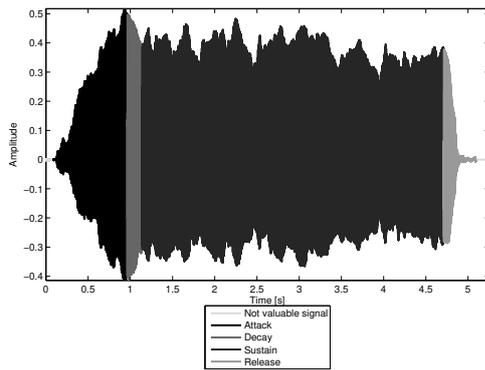


Figure 1. Example of a sung phrase divided into ADSR sections

The first phase of voice production perceived from the point of view of the ADSR envelope looks highly variable as its values increase from 0 up to the highest envelope value. However, if to take into account the physiological aspects of the human body, it will turn out that the production of sound with varying loudness and stable fundamental frequency requires a lot of effort and experience. This leads to the suggestion that greater differences in the qualitative parameters between the initial phase of voice emission and the phase with stable amplitude may be more visible in people with less vocal experience - which should also be noted when analysing the obtained values.

## II. RESEARCH CONDITIONS

The database used here was created as part of the research project of the West Pomeranian University of Technology: 'Computer methods to support the process of choral voice training' quote [10] and expanded at a later time. The content of the database allows to estimate selected parameters of the sung voice. It is possible to examine, for example, the intonation [11], the function vibrato [1], tremolo, sonority, noise [12] and other variables. It is possible to carry out more general database research, such as for example the voice quality evaluation [13], [14].

For this study, recordings containing the vowel /a/ sung on one pitch for a few seconds, were selected. In the further part of the analysis, the subsection will be the initial part of the signal and the middle part (sustain) of the samples.

The recordings used in this article were made in a specially arranged environment, with appropriate conditions for the recording session. All recordings were done with a resolution of 24 bits and with a sampling rate of 48 kHz. All singers were provided with referential signal at the beginning of recording of each sample. The process was carried out under the supervision of an expert to ensure the best quality of the samples.

The analysed group of singers consists of 16 men and 7 women. All these people have so much vocal experience that they are able to sing the sound at a given frequency. The examined persons are characterized by a varied work experience

in the team (1-20 years). The pitches range recorded for each person reflects their vocal abilities - recorded sound represents the person's voice scale.

## III. RESEARCH IDEA

The aim of the research was to confirm or reject the hypothesis that that values of the quality parameters estimated for the first part of the signal present worse quality of the singing comparing to the middle part of the signal.

To reach the goal of the study a number of acoustic (vocal) parameters can be used. Some of them are: SPR, LTAS. Another popular are: jitter and shimmer measures, harmonic-to-noise-ratio (HNR), formants (including singer's formant (SF)), Spectra Centroid, energy ratio (ER), percentual variability (PV) [15] and others.

Many of those mentioned above are used in this study. Analysing acoustic parameters we were observing a differences in values estimated for the Entry of phonation and the middle.

The chosen acoustic parameters have been estimated and analysed based on the recorded vowel /a/. The set of the estimated parameters consisted of the most recognized by the scientists in the field of voice analysis:

- Jitter,
- Shimmer,
- HNR35,
- SPR.

Above were implemented and calculated using Praat [16] (via Parselmouth [17]) and Matlab using dedicated libraries (VoiceSauce [18], YIN [19]) or our own implementations.

Jitter and shimmer are the two common perturbation measures in acoustic analysis. Jitter is a measure of frequency instability, while shimmer is a measure of amplitude instability. In Praat we have access to multiple kinds of jitters (local and local absolute, RAP, PPQ5, Jitter DDP) and shimmers (local and local in dB, APQ3, APQ5 and APQ11, Shimmer DDP)

Harmonics to Noise Ratio (estimated here in Matlab via VoiceSauce [18]) indicates ratio of harmonics values comparing to noise level. In our case HNR35 is ratio measured between 0-3500Hz.

SPR (Singing Power Ratio), for the needs of this publication, was calculated in Matlab on the basis of [20].

For each sample we omit 5% of signal from both endings of file just to be sure it contains valid signal. Remaining signal was divided equally in to five parts and for each of them we calculated total set of parameters. They were named as SET 1, SET 2 and so on. For the analysis we've chosen SET 1 which we believe is the most unstable and SET 3, which we believe that it is the most stable part of the signal in the context of parameters fluctuations.

## IV. THE RESULTS

As mentioned earlier, the study consisted in determining advanced voice parameters for acquired voice samples in the initial and middle part, and then on observing general trends.

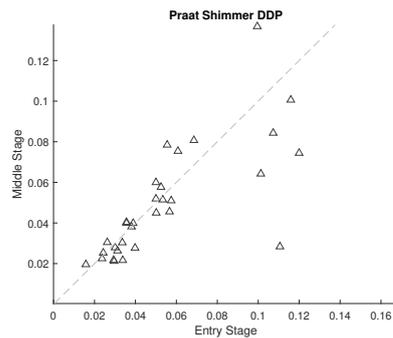


Figure 2. SET1 to SET3 Shimmer DDP relation - singer s34f

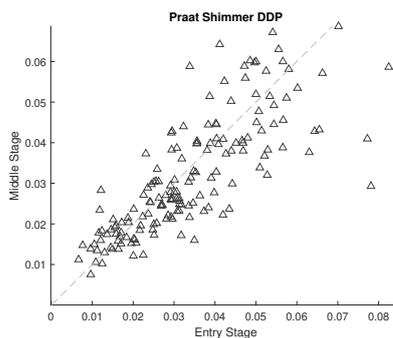


Figure 3. SET1 to SET3 Shimmer DDP relation – all females

One of the estimated parameters was the average absolute difference between consecutive differences between the amplitudes of consecutive periods (Shimmer DDP).

The shimmer DDP, similarly to other measures of this class, indicates the magnitude of changes in amplitudes of appropriate periods and can be identified with voice tremors based on small but high frequency changes in volume. This phenomenon adversely affects the quality of the voice produced. In order to check whether the SET3 presents the improvement of the DDP parameters in relation to the first set, the values from both sets were compared. In the presented figure 2, prepared for the example singer s34f, most of the Shimmer DDP values, shown in relation SET1 to SET3, are below the line determining the lack of differences/changes what can it mean, in that particular case, that the voice quality may be improved.

What is the situation for all female voices in the context of this particular parameter?

Also in the case of the Figure 3, there is a tendency to decrease values of irregularity parameters. Next, a simple statistic is presented, which consists of determining the number of 'improved' values for the analysed parameters. Table I contains the values obtained for all individuals taking part in the study.

It's shows that an improvement in quality may be observed. Particular attention should be paid to the SPR parameter, which is the ratio indicating ratio of two different formants in the Long Term Average Spectrum in the analysed signal

Table I

THE TABLE OF 'IMPROVEMENT' FOR ALL RECORDED SINGERS

	Better	Worse	Unchanged
Praat jitter (local)	666	369	16
Praat jitter (local absolute)	669	372	10
Praat jitter (rap)	579	452	20
Praat jitter (ppq5)	630	404	15
Praat Jitter DDP	579	452	20
Praat shimmer (local)	642	372	37
Praat shimmer (localdB)	654	357	40
Praat shimmer (apq3)	621	407	23
Praat shimmer (apq5)	637	371	40
Praat shimmer (apq11)	643	356	46
Praat Shimmer DDP	621	407	23
HNR35-mean	437	602	12
HNR35-std	588	439	24
SPR	627	389	26

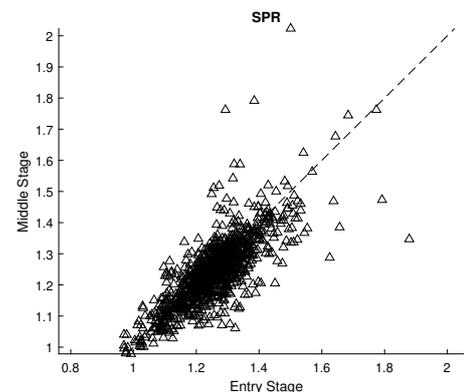


Figure 4. SPR for all singers

and is understood as an important quantitative measurement for evaluating singing voice quality for all voice types. It can be seen that this ratio improved in most of the cases. At the same time. Most of the jitter and shimmer values have been decreased, and what is more, values of the standard deviation for HNR35 have decreased. This should be considered as a simple indicator which shows that a stable part of the signal represents a signal which may have better vocal quality.

As the next, we present an 'improvement graph' for SPR values. Looking at the figure 4, it can be seen that in this case above 60% of the values has been improved. The above figures and tables show uncritical statistics and comparisons, estimated for all samples in the database. However, samples that were sung at the limit of a singers abilities should be excluded and statistics should be set again. It will also be interesting to identify changes in quality parameters only for the most comfortable samples (sung at the very centre of the vocal range).

#### A. Analysis for the comfort vocal range

An attempt was made to look at the data from which singing samples were excluded on the border of the possibilities of particular people. In the scope of sung samples, 3 samples were removed from the bottom of the range and 3 from the top of the range to analyse the range of comfortable sounds for

Table II  
IMPROVEMENT OF THE QUALITY PARAMETERS FOR THE COMFORT VOCAL RANGE

	Better	Worse	Unchanged
Praat jitter (local)	555	303	26
Praat jitter (local absolute)	537	276	71
Praat jitter (rap)	486	394	4
Praat jitter (ppq5)	521	335	28
Praat Jitter DDP	486	394	4
Praat shimmer (local)	541	311	32
Praat shimmer (localdB)	553	303	28
Praat shimmer (apq3)	521	350	13
Praat shimmer (apq5)	539	309	36
Praat shimmer (apq11)	544	296	43
Praat Shimmer DDP	521	350	13
HNR35-mean	379	499	6
HNR35-std	496	370	18
SPR	525	331	22

Table III  
IMPROVEMENT OF THE QUALITY PARAMETERS FOR THE MIDDLE VOCAL RANGE

	Better	Worse	Unchanged
Praat jitter (local)	240	113	17
Praat jitter (local absolute)	234	106	30
Praat jitter (rap)	206	158	6
Praat jitter (ppq5)	222	141	7
Praat Jitter DDP	206	158	6
Praat shimmer (local)	235	113	22
Praat shimmer (localdB)	238	113	19
Praat shimmer (apq3)	232	131	7
Praat shimmer (apq5)	228	121	21
Praat shimmer (apq11)	233	123	14
Praat Shimmer DDP	232	131	7
HNR35-mean	162	204	4
HNR35-std	221	140	9
SPR	216	143	10

singers. The answer was to find out if the presence of samples in the recordings that are not comfortable for singers will have a significant effect on the overall picture of the results obtained.

The results presented in the table II show that the overall picture of the whole has not changed - all ratios have improved mostly. In order to determine the influence of border samples on the result correction, we determined the improvement factors for both situations as the ratio of corrected values to all samples (for all parameters) and compared with the results after discarding border samples. There has been a slight improvement here, which does not mean that it does not matter. In both cases a factor of over 60% was obtained, which indicates that the extreme samples do not affect the results. This may be due to the fact that the recordings paid attention to the comfort of singing and interrupted the session at the moment when singing a certain pitch of sound made it difficult.

#### B. Analysis for the middle of vocal range

In the next scenario, samples from the middle of the vocal range were selected for the study and shown in table III. In this scenario, the results increased on average by a few percentage points. Some parameters showed improved values in 70% of cases. This shows that while recording of the samples is worth to determine the most comfortable sounds to sing and choose for testing those from the centre of the vocal ranges.

#### C. Analysis for the corresponding frequencies

In the next scenario, samples from the middle of the vocal range were selected for the study and shown in table IV. The quality coefficients aggregated for selected frequencies finally confirm the hypothesis that the central part of the recording is more valuable for the analyses.

Among many qualitative parameters estimated for the recorded samples and aggregated for particular frequencies sung by the surveyed persons, for the presentation jitter (local, absolute) was chosen. As it was mentioned before, jitter is a measure of frequency instability. Differences of the values

of that parameter gives us information about differences of the quality of the signal. In the presented example, the jitter parameter (mean value) decreased in the case of the middle segment in the case for most of the investigated sung frequencies (additionally standard deviation of the parameter also decreased) so it should be considered as the final confirmation of the hypothesis that the central part of the sample, associated with the sustain phase, presents a signal with greater stability and thus better quality.

## V. CONCLUSION

The article in general concerns the subject of signal analysis and is focused on the analysis of the quality of the signal representing the voice, in particular the voice of the singers. In the process of analysing of the voice quality, as with any signal, it is important to have samples whose content faithfully reflects the examined features to the maximum. In the case of singing samples analysis, the specifics of generating this signal should be taken into account. The singing signal characteristics can be described in an approximate way using the ADSR model. It indicates that in the initial phase of voice production and in the final phase, physiological phenomena occurs (reflected in ADSR by the attack and decay phases), which may affect the analysed features. In the analysis the middle part of the stable phase should be taken into account. The problem starts when the samples are too short. Very often it happens that the people being recorded try to shorten the sung phrase. When the analysed recording is too short, the impact of the attack phase becomes noticeable, as documented in this article. All values of the analysed signals, for the most of the samples, indicated higher signal quality in the sustain phase. This is best seen in the case of the quality coefficients aggregated for frequencies. Additionally it was confirmed that samples from the middle of the vocal range are those the best reflecting the voice of the singer.

The results concerning voice quality analysis presented here may be useful for constructing a singing quality assessment system. A large number of the results obtained for this study

Table IV  
QUALITY COEFFICIENTS AGGREGATED FOR FREQUENCIES - PART OF THE RESULTS - JITTER (LOCAL, ABSOLUTE)

Sung Frequency [Hz]	Stage	Mean value	Standard Deviation	Percentile 25%	Median	Percentile 75%
146.8324	Entry	3.2081e-05	1.1301e-05	2.2417e-05	3.0432e-05	4.2687e-05
	Middle	2.4737e-05	8.3209e-06	1.7707e-05	2.2877e-05	2.8792e-05
155.5635	Entry	2.2129e-05	7.1437e-06	1.6498e-05	2.2769e-05	2.4551e-05
	Middle	1.8161e-05	7.3445e-06	1.2832e-05	1.6569e-05	2.3387e-05
164.8138	Entry	1.8351e-05	7.3706e-06	1.2684e-05	1.8483e-05	2.1523e-05
	Middle	1.5216e-05	7.1313e-06	1.0105e-05	1.2216e-05	1.8996e-05
174.6141	Entry	1.7062e-05	7.4216e-06	1.1709e-05	1.4635e-05	2.1104e-05
	Middle	1.5071e-05	7.8123e-06	9.859e-06	1.2466e-05	1.6579e-05
184.9972	Entry	1.3369e-05	6.2646e-06	9.6237e-06	1.1677e-05	1.4684e-05
	Middle	1.1412e-05	4.0202e-06	8.406e-06	9.9462e-06	1.4236e-05
195.9977	Entry	1.1873e-05	4.703e-06	8.4799e-06	1.3474e-05	1.5916e-05
	Middle	1.0632e-05	6.1637e-06	6.4303e-06	8.584e-06	1.2656e-05
207.6523	Entry	7.6085e-06	3.5426e-06	5.4417e-06	6.8487e-06	8.2179e-06
	Middle	6.6576e-06	4.2819e-06	3.6837e-06	5.8339e-06	7.4392e-06
220	Entry	5.8739e-06	1.5366e-06	4.6282e-06	6.1168e-06	7.1197e-06
	Middle	4.9587e-06	1.6141e-06	3.6099e-06	5.1528e-06	6.3075e-06
233.0819	Entry	6.653e-06	4.0259e-06	4.6118e-06	5.1807e-06	6.7364e-06
	Middle	5.5737e-06	3.6728e-06	3.6742e-06	4.5198e-06	5.4754e-06
246.9417	Entry	6.9293e-06	4.3558e-06	4.9796e-06	5.2084e-06	6.693e-06
	Middle	7.3143e-06	5.343e-06	3.0391e-06	6.4562e-06	8.0467e-06
261.6256	Entry	7.9658e-06	3.076e-06	6.03e-06	8.0923e-06	9.9593e-06
	Middle	6.3619e-06	4.8693e-06	3.9123e-06	4.3487e-06	6.9787e-06
277.1826	Entry	5.3634e-06	1.7809e-06	4.233e-06	5.465e-06	6.4937e-06
	Middle	6.2812e-06	3.2574e-06	3.6387e-06	6.5349e-06	8.9237e-06
293.6648	Entry	4.6625e-06	1.4611e-06	3.6059e-06	4.4312e-06	5.7768e-06
	Middle	7.5257e-06	5.2186e-06	3.7047e-06	6.9339e-06	1.1495e-05

requires further, deeper analysis and may lead to subsequent applications.

#### ACKNOWLEDGMENT

The authors would like to thank the members of the Jan Szyrocki Memorial Choir from the West Pomeranian University of Technology who have devoted their time to recording voice samples. Thanks are also due to the experts in the field of voice emission, especially for Mr. Adam Kuliś from Study of Culture, for substantive consultations.

#### REFERENCES

- [1] E. Pórolniczak and M. Kramarczyk, "Analysis of the signal of singing using the vibrato parameter in the context of choir singers," *Journal of Electronic Science and Technology*, vol. 11, no. 4, pp. 417–423, December 2013.
- [2] J. Godino-Llorente, P. Gomez-Vilda, and M. Blanco-Velasco, "Dimensionality reduction of a pathological voice quality assessment system based on gaussian mixture models and short-term cepstral parameters," *Biomedical Engineering, IEEE Transactions on*, vol. 53, no. 10, pp. 1943–1953, Oct 2006. doi: 10.1109/TBME.2006.871883
- [3] K. Omori, A. Kacker, L. M. Carroll, W. D. Riley, and S. M. Blaugrund, "Singing power ratio: Quantitative evaluation of singing voice quality," *Journal of Voice*, vol. 10, no. 3, pp. 228 – 235, 1996. doi: [http://dx.doi.org/10.1016/S0892-1997\(96\)80003-8](http://dx.doi.org/10.1016/S0892-1997(96)80003-8). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0892199796800038>
- [4] Y. Meron and K. Hirose, "Separation of singing and piano sounds." in *ICSLP*, 1998.
- [5] M. Muller, D. P. Ellis, A. Klapuri, and G. Richard, "Signal processing for music analysis," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, no. 6, pp. 1088–1110, 2011.
- [6] A. Holzapfel, Y. Stylianou, A. C. Gedik, and B. Bozkurt, "Three dimensions of pitched instrument onset detection," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 18, no. 6, pp. 1517–1527, 2010.
- [7] L. Mazurowski, "Computer models for algorithmic music composition," in *Computer Science and Information Systems (FedCSIS), 2012 Federated Conference on*. IEEE, 2012, pp. 733–737.
- [8] K. Jensen, "Envelope model of isolated musical sounds," in *Proceedings of the 2nd COST G-6 Workshop on Digital Audio Effects (DAFx99)*, 1999.
- [9] R. M. Alderson, *Complete handbook of voice training*. Parker Publishing Company, 1979.
- [10] M. Łazoryszczak and E. Pórolniczak, "Audio database for the assessment of singing voice quality of choir members," *Elektronika: konstrukcje, technologie, zastosowania*, vol. 54, no. 3, pp. 92–96, 2013.
- [11] E. Pórolniczak and M. Łazoryszczak, "Quality assessment of intonation of choir singers using f0 and trend lines for singing sequence," *Metody Informatyki Stosowanej*, pp. 259–268, 2011.
- [12] E. Pórolniczak and M. Kramarczyk, "Computer analysis of the noise component in the singing voice for assessing the quality of singing," *Przegląd Elektrotechniczny*, vol. 91, pp. 79–83, 2015.
- [13] E. Pórolniczak and M. Kramarczyk, "Formant analysis in assessment of the quality of choral singers," in *Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), 2013*, Sept 2013. ISSN 2326-0262 pp. 200–204.
- [14] P. Zwan and B. Kostek, "System for automatic singing voice recognition," *Journal of the Audio Engineering Society*, vol. 56, no. 9, pp. 710–723, 2008.
- [15] E. H. Buder, "Acoustic analysis of voice quality: A tabulation of algorithms 1902–1990," *Voice quality measurement*, pp. 119–244, 2000.
- [16] P. Boersma and D. Weenink, "Praat: doing phonetics by computer [Computer program]," Version 6.0.43, retrieved 8 September 2018 <http://www.praat.org/>, 2018.
- [17] Y. Jadoul, B. Thompson, and B. de Boer, "Introducing Parselmouth: A Python interface to Praat," *Journal of Phonetics*, vol. 71, pp. 1–15, 2018. doi: <https://doi.org/10.1016/j.wocn.2018.07.001>
- [18] Y.-L. Shue, P. Keating, C. Vicens, and K. Yu, "Voicesauce," *p. Program available online at http://www.seas.ucla.edu/spapl/voicesauce/*. UCLA, 2009.
- [19] A. De Cheveigné and H. Kawahara, "Yin, a fundamental frequency estimator for speech and music," *The Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1917–1930, 2002.
- [20] E. Yumoto, W. J. Gould, and T. Baer, "Harmonics-to-noise ratio as an index of the degree of hoarseness," *The Journal of the Acoustical Society of America*, vol. 71, no. 6, pp. 1544–1550, 1982.



# Automatic Assessment of Narrative Answers Using Information Retrieval Techniques

Liana Stanescu

University of Craiova, Faculty of Automation,  
Computers and Electronics, Craiova, Romania  
Email: stanescu@software.ucv.ro

Beniamin Savu

University of Craiova, Faculty of Automation,  
Computers and Electronics, Craiova, Romania  
Email: benisavu@gmail.com

**Abstract**—In this paper we propose a novel system for automatic assessment of narrative answers using information retrieval algorithms. It is designed to help professors to evaluate the answers that they receive from their students. It is a Java application that communicates through a REST API. This REST API has at its core the Lucene library and exposes all the great functionalities that Lucene has. The application has one UI for the students and one UI for the professor. The student will select the professor, select the question, upload the answer and send it. The professor will evaluate the student answer using the algorithms that will be discussed in this paper. Also in this paper a series of experiments will be presented, and their result will give us a better understanding of the algorithms and have a taste of how they work.

## I. INTRODUCTION

IN our days e-learning is becoming more and more popular because of the benefits that it can offer. Because evaluating students on-line represent an important action, research in the domain has been focused on improving the on-line assessment systems, by integrating various methods for accomplishing this desire. The majority of online assessment systems have integrated numerous types of questions that can be evaluated and graded easily, based on direct matching: true/false, multiple choice, fill in the blank.

However, there is the general opinion that these types of objective tests are not enough. There are many topics, especially in the domain of human science, as well as technical domain, where the evaluation of a student cannot be complete without narrative answers. In this case, the student must formulate the answer to questions in the form of free text. Thus, it is desired for an online assessment system to integrate assessment of objective questions and narrative questions together for a complete evaluation of the students' capacity of assimilating information.

Nowadays, there is no viable method that can evaluate the answers given by the students. Having the computer understand our language and not only numbers will bring great benefit because they can process faster than us and come up with solutions in seconds rather than hours, days,

weeks or months. So we will give the computer a set of answers and compare it with the model answer which is the correct answer.

In our original work, the comparison between the student answer and the correct answer will be done by applying some information retrieval algorithms. Each algorithm will be taken one by one to investigate on how it works, how it compares to the other algorithms and what results it gives to the test data. The algorithms that will be used are Vector Space Model (VSM), Bigram and Language Analyser.

## II. RELATED WORK

In the last years, there has been research in the domain of automatic assessment of narrative answers. The results have been integrated in certain academic or commercial platforms [4], [5], [6], [7], [8], [9], [10].

A classification of these techniques can be found in [4]:

- Statistical methods
- Text Categorization Techniques (TCT)
- Information Retrieval algorithms.
- Full Natural Language Processing (NLP)
- Clustering
- Hybrid approaches that combine several techniques

Although the techniques may seem very different, the general idea that stands at the base of these systems is the same: to compare the student's answer (or candidate answer) with the teacher's ideal answer (or reference answer). The closer they are, the higher the student's score is.

There is a series of studies on the use of Information Retrieval Algorithms in automatic evaluation of free answers that indicates their efficiency in the domain. In [11] the author presents a comparative study of 5 algorithms. The model answers and student's answers are represented as vectors and then similarities between them computed by using cosine similarity. The obtained results are very satisfying. In [12] is also presented a comparison of 3 algorithms: Fingerprint, winnowing algorithms and the cosine similarity that are widely used to compare documents.

### III. SYSTEM OVERVIEW

The system has a modular architecture [2] presented in the figure 1. The user interacts with the system through the Java Application which has access to the database and to the REST API. The application modules communicate through Data Transfer Object (DTO).

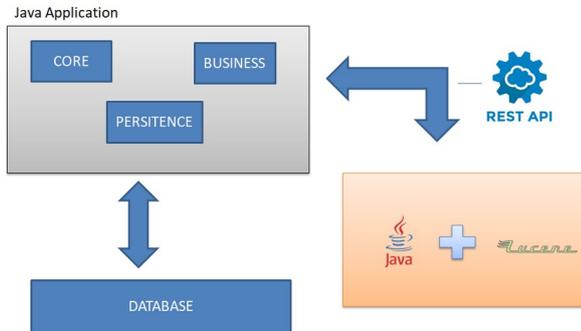


Fig. 1. Block Diagram of System Modules

A DTO is an object that carries data between processes. In other words, DTOs are simple objects that should not contain any business logic but may contain serialization and deserialization mechanisms for transferring data [3].

#### A. System Features

When the application is run the user will be prompted with a login screen and an option to register. The user will have to create a new account by selecting that option and fill up the register form.

In this system a user can have two roles. He can be a professor or a student. The application will provide to the student a combobox with the list of professors that are available. The student must select one of the professors. The selected professor will receive his answers. After this the student can choose whether to upload a file which contains the answer or insert the answer in the text area provided by the application. The student can edit his answer as long as the answer has not been send to the professor.

Once the answer is complete and the professor selected the student can now safely send the answer. The answer will be saved in the database and the professor that was selected will receive it for evaluation.

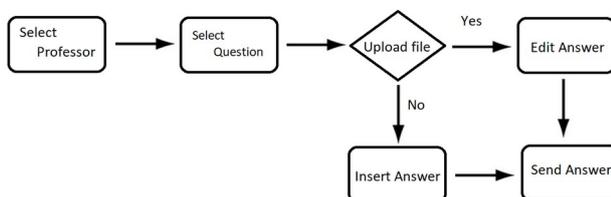


Fig. 2. Student activity diagram

The application will provide the professor with all the answers that the students submitted to him. The answers will come with the status of NOT\_PROCESSED. The professor will now choose a reference text to evaluate the answers that

were submitted. To evaluate the answers, the application will use different algorithms. But before the professor can evaluate, he must index all of the answers.

After the process of indexing is done the status for the answers will change from NOT\_PROCESSED to INDEXED and he will be able to apply all the algorithms. The evaluation is done using the three algorithms: VSM, Bigram and Language Analyser. For each algorithm there will be a separate score. The higher the score, the closer it is to the correct answer entered by the professor.

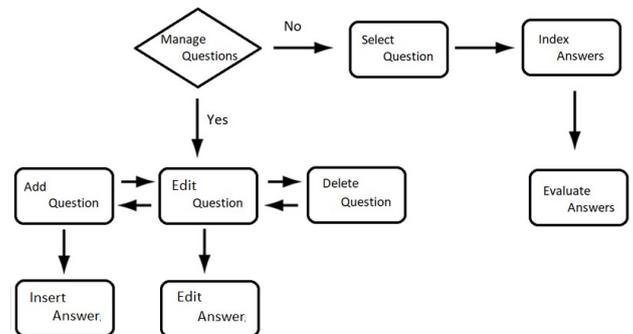


Fig. 3. Professor activity diagram

These are the high-level functionalities for the student and the professor, how they interact, what are the flows for them. Next we will take a closer look on how the index process works, how the algorithms work and all that is happening behind the scenes.

#### B. Inverted Index

Lucene uses a structure called an inverted index, which is designed to allow very fast full-text searches [13].

An inverted index consists of a list of all the unique words that appear in any document, and for each word, a list of the documents in which it appears.

To create an inverted index, we first split the content field of each document into separate words (which we call terms, or tokens), create a sorted list of all the unique terms, and then list in which document each term appears.

### IV. ALGORITHMS

The algorithms are the heart of the system. Let's have a closer look at those algorithms and see how they actually work.

#### A. Vector Space Model

Once we have a list of matching documents, they need to be ranked by relevance. Not all documents will contain all the terms, and some terms are more important than others. The relevance score of the whole document depends (in part) on the weight of each query term that appears in that document. The weight of a term is determined by three factors [1]: term frequency, inverse document frequency and field-length norm.

### B. Bigram

When words are used in conjunction with each other, they express an idea that is bigger or more meaningful than each word in isolation. The two clauses “I’m not happy I’m working” and “I’m happy I’m not working” contain the same words, in close proximity, but have quite different meanings [1]. If, instead of indexing each word independently, we were to index pairs of words, then we could retain more of the context in which the words were used. These word pairs (or bigrams) are known as shingles [1]. Of course, shingles are useful only if the user enters the query in the same order as in the original document. But this point is an important one: it is not enough to index just bigrams; we still need unigrams, but we can use matching bigrams as a signal to increase the relevance score [1].

Not only shingles are more flexible than phrase queries, but they perform better as well.

### C. Language Analyser

Full-text search is a battle between precision, returning as few irrelevant documents as possible, and recall, returning as many relevant documents as possible. While matching only the exact words that the user has queried would be precise, it is not enough. We would miss out on many documents that the user would consider to be relevant. Instead, we need to spread the net wider, to also search for words that are not exactly the same as the original but are related [1]. There are several lines of attack: rRemove diacritics like ‘, ^, and “ so that a search for rôle will also match role, and vice versa; remove the distinction between singular and plural; remove commonly used words or stop words to improve search performance; Including synonyms; check for misspellings or alternate spellings, or match on homophones—words that sound the same.

Lucene ships with a collection of language analyzers that provide good, basic, out-of-the-box support for many of the world’s most common languages [13].

## V. EXPERIMENTS AND RESULTS

We needed some datasets to test the algorithms, so for the first dataset we asked our friends to give 15 answers regarding any topic that they want. They chose “sadness”. The following text is the relevance answer: „Sadness is a normal feeling, an emotion we occasionally feel and should not be denied. Sadness is necessary; otherwise we could not appreciate the beautiful moments. It's normal not to feel good when you suffer a loss, when you're disappointed when something goes wrong.”

### A. Experiment 1

Answer: **Sadness is an emotion we sometime are feeling and should not deny it. Without those moments we will not be able to appreciate the value of the world. This pain feels like a push of the soul, it is like a hollow in the soul**

that does not let you hope for better. **It's normal to feel like this, to be disappointed when something goes wrong.**

Score: VSM(18.4), Bigram(29.3), Language Analyzer(31.6)

We see that the answer contains quite a few words compared to the relevance text. Bigram scores very high because of the two word pairing that appear in the relevance text as well as in the answer text („Sadness is”, „an emotion”, „it's normal”). Language Analyzer also gives us a high score because of the stemming function that it applies to the our texts, so words like “feeling”, “feels” are transformed to “feel”, “sadness” to “sad” and so on. Basically every word from the texts are stemmed to their root form, thus increasing the chances to find the same word multiple times.

### B. Experiment 2

Answer: **Sadness is an emotional pain we often try to hide from others. We experience beautiful moments when we are not sad. It's normal not to feel good when you suffer a loss, when you're disappointed when something goes wrong.**

Score: VSM(26.5), Bigram(52.8), Language Analyzer(47.1)

We see that this answer contains an entire phrase that also appears in the relevance text and every algorithm gave us high scores. Here we see bigram scoring very big. By having the words in the exact order as in the relevance text, bigram scores higher, informing us that this answer is very relevant.

We will not show the results for the entire dataset. We chose only the ones that are the most interesting. The other results are similar with the ones presented above, either they are very close or they are very far or they are somewhere in the middle.

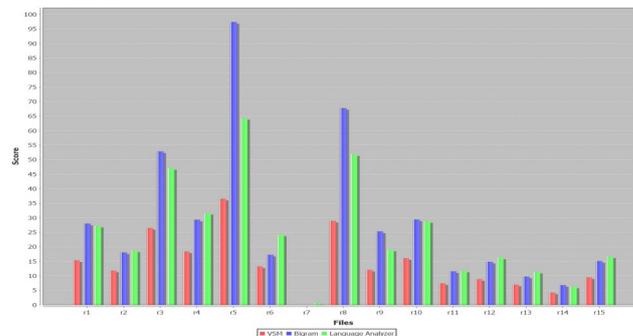


Fig. 4. Bar chart result for the first dataset

RED: VSM; BLUE: Bigram; GREEN: Language Analyser

As you can see in figure 4, for the first dataset, Bigram and the Language Analyser both perform better than VSM. This is because VSM takes into account only the words that are found in the answers, it does not perform any additional operations. Bigram searches through the answers not only with single word queries like VSM but also with pairs of 2 words, thus keeping some of the semantic for the respective

answer. Language Analyser besides using the single word queries also uses the stemming of the words, so any variation of that word will be taken into consideration. Bigram and Language Analyzer are the way to go. Each one has their strengths and weaknesses. So on our dataset, for some answers bigram scores the most and for other answers Language Analyzer scores the most.

We will now look at the physics dataset. We will present the bar chart and the relevance order that the answers have for each algorithm (figure 5).

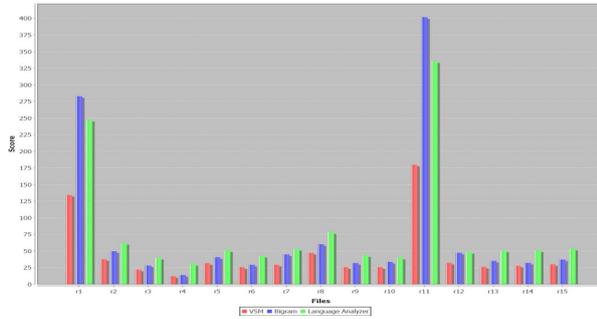


Fig. 5. Bar chart result for the second dataset

RED: VSM; BLUE: Bigram; GREEN: Language Analyser

First are VSM and Bigram. There is a big difference in the relevance order between them. By having a larger answer and a larger relevance text those results are normal for bigram, because bigram has more chances to find pairs of 2 words, thus making the answers more relevant.

Now let's look at Bigram and Language Analyzer. There are also differences in the relevance order between those two. Again this has to do with the fact that we have bigger answers and a bigger relevance text. So, the Language Analyzer is taking full advantage of its stemming operation, because a larger text means that the Language Analyzer will have more words that have the same root, thus a higher score.

The mean for each algorithm calculated from the data acquired in the 2 datasets appears in figure 6.

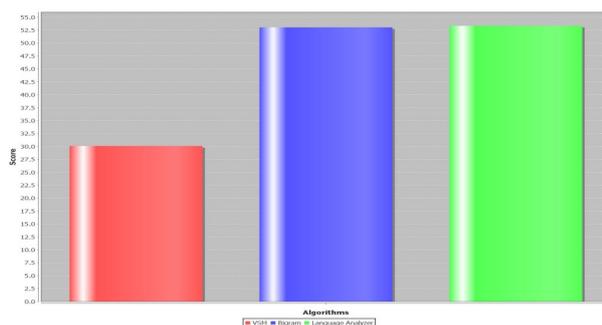


Fig. 6. Algorithms mean result

RED: VSM; BLUE: Bigram; GREEN: Language Analyser

We have the following results: VSM Mean: 30.1, Bigram Mean 53.05, Language Analyser Mean: 53.35.

As it can be seen the difference between Bigram and Language Analyser is very slim, advantage to Language

Analyser. Either one of the two can be used to automatically evaluate answers.

## VI. CONCLUSIONS

The paper presents our novel system that automatically evaluates narrative answers using information retrieval techniques. As seen in the experiments above, Bigram and Language Analyser were the algorithms that performed the best, with Language Analyser having a small advantage over Bigram. But nonetheless both of them are very good and have scored good results on our test datasets.

Even though the objective was achieved it is not yet complete. Of course we can give scores to our answers, but we want some kind of mechanism to interpret this score and give that answer a grade. Also we might look to improve the algorithms to perform even better, a solution might be to combine Bigram and Language analyzer concepts. The performed experiments have shown satisfying results. For the time being, the module is integrated in an e-learning platform, in order to further be used and evaluated by the professors.

## REFERENCES

- [1] C. Gormley and Z. Tong, Elasticsearch: The Definitive Guide available via web at <https://www.elastic.co/guide/en/elasticsearch/guide/current/index.html>
- [2] G. Wielenga, "How to Split an Application into Modules?" <https://dzone.com/articles/how-to-split-into-modules>, 2009
- [3] [https://en.wikipedia.org/wiki/Data\\_transfer\\_object](https://en.wikipedia.org/wiki/Data_transfer_object)
- [4] D. Perez, "Automatic evaluation of users' short essays by using statistical and shallow natural language processing techniques", Retrieved from <https://pdfs.semanticscholar.org/025e/cb63d3322608e8f3073965ee9a0fc4d51e63.pdf>, 2004
- [5] D. Perez, E. Alfonseca and P. Rodriguez, "Adapting the automatic assessment of free-text answers to the students profiles", Retrieved from [https://dspace.lboro.ac.uk/dspace-jspui/bitstream/2134/2000/1/PerezD\\_AlfonsecaE.pdf](https://dspace.lboro.ac.uk/dspace-jspui/bitstream/2134/2000/1/PerezD_AlfonsecaE.pdf), 2005
- [6] D. Perez, O. Postolache, E. Alfonseca, D. Cristea, and P. Rodriguez, "About the effects of using Anaphora Resolution in assessing free-text student answer", in *Proceedings of Recent Advances in Natural Language Processing Conf.* Borovets, Bulgaria, pp.380-386, 2005.
- [7] T. Mitchell, T. Russell, P. Broomhead and N. Aldridge, "Towards robust computerised marking of free-text responses", Retrieved from [https://dspace.lboro.ac.uk/dspace-jspui/bitstream/2134/1884/1/Mitchell\\_t1.pdf](https://dspace.lboro.ac.uk/dspace-jspui/bitstream/2134/1884/1/Mitchell_t1.pdf), 2002
- [8] J. Burstein, C. Leacock and R. Swartz, "Automated evaluation of essays and short answers", in *Proceedings of the International CAA Conference.* Loughborough: Loughborough University Davis, 2001
- [9] E. Alfonseca and D. Perez, "Automatic assessment of short questions with a bleu-inspired algorithm and shallow nlp", *Advances in Natural Language Processing*, Springer Verlag, pp. 25-35, 2004
- [10] D. Pérez-Marín, I. Pascual-Nieto and P. Rodríguez, "Computer-assisted assessment of free-text answers", *Knowledge Eng. Review*, vol. 24, no. 4, pp.353-374, 2009
- [11] M. M. Hassan, "Experiments in Automatic Assessment Using Basic Information Retrieval Techniques", *Knowledge, Information and Creativity Support Systems*, Berlin, Springer, pp.13-21, 2011
- [12] K. T. Tung, N. D. Hung and L. T. M. Hanh, "A Comparison of Algorithms used to measure the Similarity between two documents", *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, vol. 4, no. 4, pp. 1117-1121, 2015
- [13] <https://lucene.apache.org/>



The proposed DC Microgrid is connected to a 10-kW PV array through an unidirectional DC-DC boost average converter, capable of generating a regulated 300V DC Bus that supplies unidirectional and bidirectional DC-DC converters that control the charging and discharging of two storage power sources, Li-ion battery and Supercapacitor, and also a 2 HP 1750 rpm PMDCM - DC load.

### III. FUZZY LOGIC PID SPEED CONTROL DESIGN AND MATLAB SIMULATIONS

In this research paper the FLC is conceived as a multi-input single-output (MISO) subsystem with two inputs and one single output. The inputs of the fuzzy controller are two measured variables whose values are collected by a set of sensors integrated into a data acquisition system describing the speed error (“e”) and the rate of change of the speed error (“ce”),  $ce = \frac{\Delta e}{\Delta t}$ , from PMDCM, the choice suggested in

[5]-[6]. The both inputs are then "fuzzified" using membership functions provided by an expert operator to determine the degree of membership in each input class. The resulting "fuzzy inputs" are evaluated using a linguistic rule base and fuzzy logic operations (AND, OR, NOT) to yield an appropriate fuzzy output and an associated degree of membership [7]-[8]. The "fuzzy output" of FLC is then “defuzzified” using a “centroid” gravity method to give a crisp output response to control the input voltage of a controlled source voltage whose output provides the regulated voltage to the PMDCM armature [8].

The fuzzy logic controller (FLC) SIMULINK model it is shown in Fig. 2.

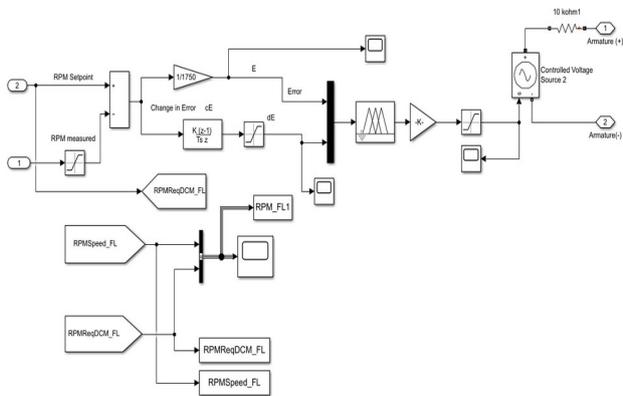


Fig.2 The Fuzzy Logic controller SIMULINK model

As it is shown in Fig.2, the inputs of the FLC, i.e. the RPM PMDCM speed error (E) and the rate of its change (cE), are converted into fuzzy linguistic variables. They are divided into a finer fuzzy partition with seven terms as it is suggested in [5]-[7], namely negative big (NB), negative medium (NM), negative small (NS), zero (ZO), positive small (PS), positive medium (PM) and positive big (PB). The range of the both inputs is [-1750 1750], and for output voltage is [-300 +300]. The universe of discourse of error,

rate of error and output is normalized to [-1, 1]. The Fuzzy Logic Designer block used to implement the SIMULINK FLC model specified in Fig. 2 it is shown in Fig. 3, where the membership functions for all two inputs and the output are represented with triangle-shaped function that it is shown in Fig. 4.

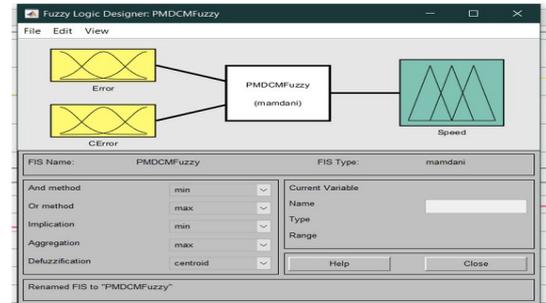


Fig.3 The Fuzzy Logic Designer block

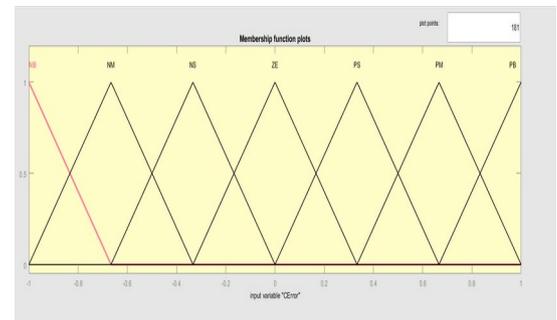


Fig.4 The membership function used for both inputs and the output of Fuzzy Logic controller

The optimal performance of the FLC design it is well depicted in 3-dimensional space of both FLC inputs and its output by a surface view shown in Fig. 5. The MATLAB simulation results of RPM FLC PMDCM speed control for a step response are shown in Fig. 6, and its robustness to changes in tracking setpoint and load torque it is shown in Fig.7.

In Fig.6 it is revealed a great performance for FLC RPM speed, a very short rising and settling time, and also a very good tracking accuracy.

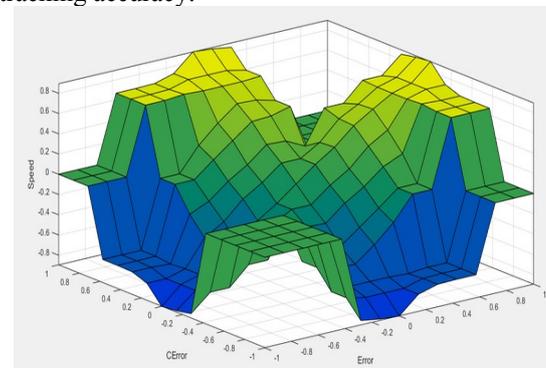


Fig.5 The surface view of the optimal performance of the Fuzzy Logic controller

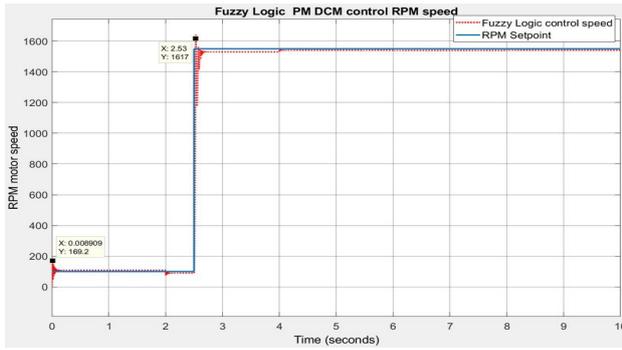


Fig.6 The FLC performance of RPM control speed versus PMDCM RPM tracking setpoint

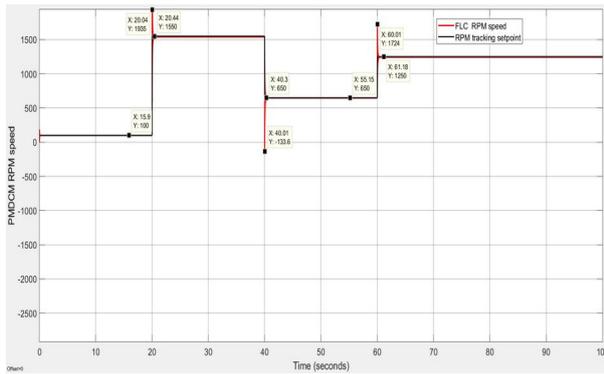


Fig.7 The FLC performance of RPM control speed versus the changes in PMDCM RPM tracking setpoint

In [9] it is suggested a combination of PID and FLC in a new structure FLPID, as it is shown in Fig.8.

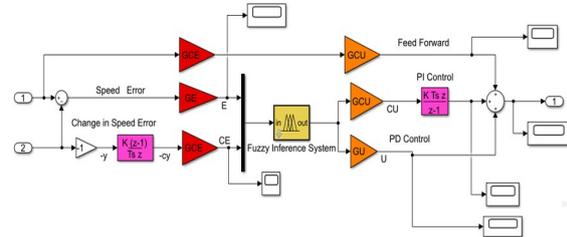


Fig.8 The Fuzzy Logic PID controller SIMULINK model (see [9])

As it is suggested in [9] the scaling factors of Fuzzy Logic PID that it is shown in Fig.8 are calculated based on PID tuned optimal values of control parameters,  $k_p, k_i, k_d$  using the advanced SIMULINK PID block option. Thus, the PID are set up for the following optimal values of parameters:

$$k_p = -2.047, k_i = -14.7, k_d = -0.0487, T_s = 0.1 \text{ that lead}$$

to the following values of the scaling factors [8]:

$$GE = 1750, GCE = GE * (k_p - \sqrt{k_p^2 - 4 * k_i * k_d}) / 2 / k_i = 190.4049$$

$$GCU = k_i / GE = -0.0084, GU = k_d / GCE = -2.5577e-04, K=1$$

The overall performance of the Fuzzy Logic PID controller it is shown in Fig. 9.

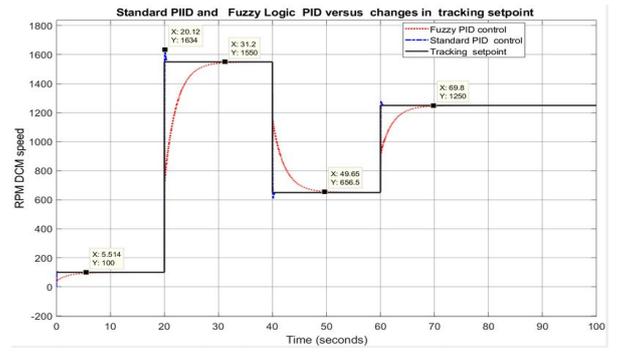


Fig.9 The Fuzzy Logic PID RPM control speed performance versus standard PID control

#### IV. PMDCM SPEED CONTROL STRATEGIES PERFORMANCE ANALYSIS

A comparison of MATLAB simulations results for the proposed PMDCM speed (RPM) control strategies, i.e. PID, FLC and their combination FLCPID, in terms of tracking accuracy, rise and settling time responses, as well as robustness to changes in tracking reference value and load torque are shown in Fig.10.

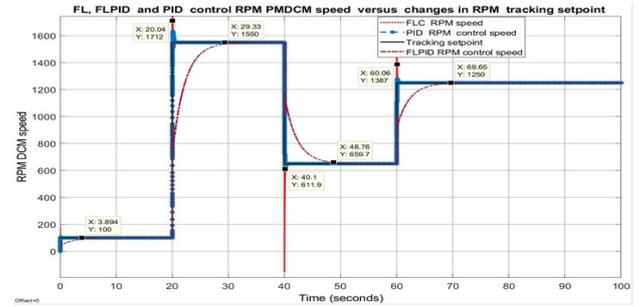


Fig.10 The PID and FLC performance of speed (RPM) control versus the changes in PMDCM RPM tracking setpoint

In the following figures Fig.11, Fig.12, and Fig. 13 there are shown the PMDCM armature current, armature voltage and the load torque profile for the combined speed (RPM) control Fuzzy Logic PID, more smooth compared to those obtained for PID and FLC separately. Also, a rigorous analysis of the performance in terms of tracking accuracy, rise time and settling time responses, and the robustness to changes in tracking setpoint and load torque indicates that PID and FLC control strategies perform slightly better compared to FLCPID, but the last one acts much smooth during the transient than first two control strategies, more useful for such of kind of applications.

#### V. CONCLUSION

In this research paper we developed a 10-kW Microgrid of 300V DC bus connected to a PV array that supplies power to a 2 HP 1750 rpm PMDCM, two storage power sources such as a Li-Ion battery and a Supercapacitor. The PV array is connected to DC Microgrid thru a DC-DC boost average

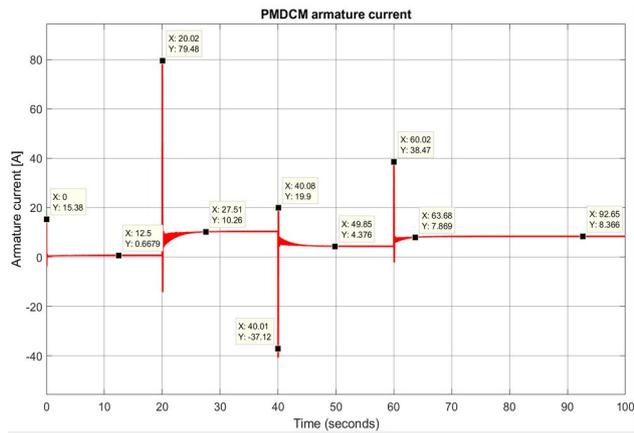


Fig.11 PMDCM armature current

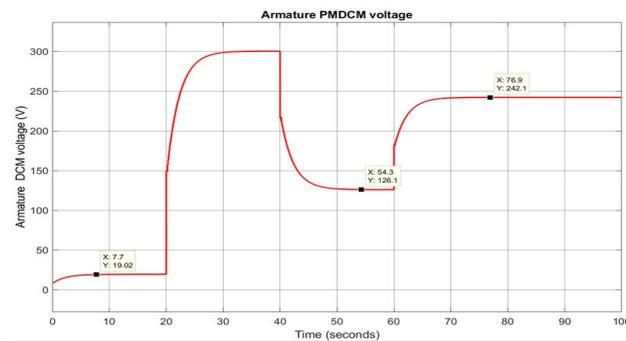


Fig.12 PMDCM armature voltage

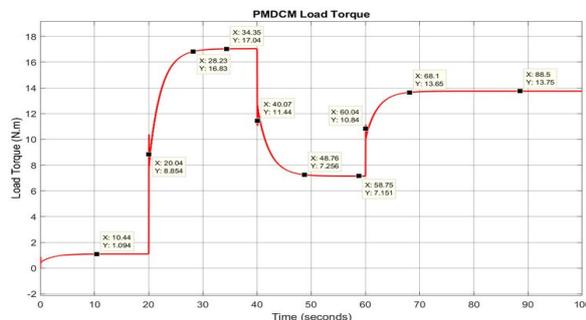


Fig.13 PMDCM load torque

converter controlled by an MPPT implemented by using the simplest PandO technique. The PMDCM is connected as a DC load thru a monodirectional DC-DC boost converter, and the both storage power sources thru two bidirectional DC-DC boost-buck converters. The investigation it was focused to develop for the proposed PMDCM a Fuzzy Logic speed control approach, and also a hybrid combination of the both control strategies, known as Fuzzy Logic PID controller.

The effectiveness of the proposed control strategies it was proved by intensive MATLAB simulations conducted on a MATLAB 2019a platform and SIMULINK. For future work we will be focused to extend the implementation of FLC and of the hybrid FLCPID control structure on a variable speed airflow fans used in a single zone or multi-zone VAV HVAC applications.

## REFERENCES

- [1] Shweta Dikshit, "Solar Photovoltaic generator with MPPT and battery Storage", International Journal of Electrical Engineering and Technology, vol.8 (3), 2017, pp. 42-49, ISSN Print: 0976-6545, ISSN online: 0976-6553.
- [2] R-E Tudoroiu, W. Kecs, M. Dobritoiu, N. Ilias, S-V Casavela, N. Tudoroiu, "Real-Time Implementation of DC Servomotor Actuator with Unknown Uncertainty using a Sliding Mode Observer", ACSIS, vol.8, pp.841-848, DOI: 10.15439/2016F95, Poland, 2016.
- [3] E-R. Tudoroiu, M. Zaheeruddin, N. Tudoroiu, D.D. Burdescu, "MATLAB Implementation of an Adaptive Neuro-Fuzzy Modeling Approach applied on Nonlinear Dynamic Systems – a Case Study", Proceedings of the Federated Conference on Computer Science and Information Systems, pp. 577–583, September 2018, ISSN: 2300-5963, ACSIS, vol.15, DOI: 10.15439/2018F38.
- [4] K. F. Hussein, "Hybrid Fuzzy-PID controller for Buck-Boost Converter in solar energy-Battery systems", Master's Thesis, Western Michigan University, May 2015.
- [5] Huang Jiang, Wang Jie, Fang Hui, "An anti-windup self-tuning fuzzy PID controller for speed control of brushless DC motor", Automatika, 2017, vol. 58 (3), pp. 321-335, DOI: 10.1080/00051144.2018.1423724.
- [6] Chuen Chien Lee, "Fuzzy Logic in control systems: Fuzzy Logic Controller-Part 1", IEEE Transactions on Systems, Man, and Cybernetics, vol.20 (2), March/April 1990, pp. 404-418.
- [7] Md. Akram Ahmad, Pankaj Rai, Anita Mahato, Megha Mahapatra, "Speed control of a DC Motor using Fuzzy Logic application", International Journal of Research in Engineering, Technology and Science, vol.7, Feb. 2017, ISSN 2454-1915, pp.1-12.
- [8] Umesh Kumar Bansal, Rakesh Narvey, "Speed Control of DC Motor Using Fuzzy PID Controller", Advance in Electronic and Electric Engineering, ISSN 2231-1297, vol. 3(9), 2013, pp. 1209-1220.
- [9] MathWorks Documentation, MATLAB R2019a, Fuzzy Logic Toolbox Examples/Implement Fuzzy PID Controller in Simulink Using Lookup Table - SIMULINK Library MATLAB R2019a.

# Object detection in the police surveillance scenario

Artur Wilkowski, Włodzimierz Kasprzak, Maciej Stefańczyk  
Insitute of Control and Computation Engineering,  
Warsaw University of Technology  
ul. Nowowiejska 15/19, 00-665 Warsaw, Poland  
Email: artur.wilkowski@pw.edu.pl  
{W.Kasprzak,M.Stefanczyk}@elka.pw.edu.pl

**Abstract**—Police and various security services use video analysis when investigating criminal activity. One typical scenario is the selection of object in image sequence and search for similar objects in other images. Algorithms supporting this scenario must reconcile several seemingly contradicting factors: training and detection speed, detection reliability and learning from sparse data. In the system that we propose a combined SVM/Cascade detector is used for both speed and detection reliability. In addition, object tracking and background-foreground separation algorithm together with sample synthesis is used to collect rich training data. Experiments show that the system is effective, useful and suitable for selected tasks of police surveillance.

## I. INTRODUCTION

**P**OLICE and various security services use video analysis when investigating criminal activity. Long surveillance videos are increasingly searched by dedicated image analysis software to detect criminal events, to store them and to initiate proper security actions (see e.g. the P-REACT project [1]). Solutions to automatic analysis of surveillance videos seem already to be mature enough, as the research community is recently also involved in major benchmark initiatives [2], [3]. The computer vision research focus is now shifted to the analysis of video data coming from handheld, body-worn and dashboard cameras and on the integration of such analysis results with police- and public-databases.

In typical object detection scenarios, there are much data to learn from and major objective is to use them in effective manner. In a security-oriented environment the user interaction should be kept as simple as possible and preferably limited only to marking single object in a selected image frame and initiating search to find occurrences of similar objects in other frames of the processed sequence or other sequences. This imposes several constraints on the Machine Vision solution that need to be addressed.

First of all the system should learn on-line or nearly on-line. Secondly - the system must perform per-frame detection quickly and provide approximate results in short time. And thirdly - to system must be able to learn from sparse data.

In this paper, an effective and time-efficient algorithm for instance search and detection in images from handheld video cameras is proposed. The system uses a discriminative approach to differentiate the object from its foreground. In order

This research was funded by NCBiR Agency, Warsaw, under BLOWIZ project, grant number DOB-BIO7/18/02/2015. The manuscript preparation was supported by statutory funds of the author's home institution (WUT).

to do so a combined Haar-Cascade detector and SVM classifier are used. We argue that this provides a very attractive trade-off between detection quality and training/detection times. Both the positive as well as negative samples are extracted only from training images.

Comparable detector solutions based on CNNs provide excellent detection performance [4]. Such solutions, however, rely on off-line training and training/detection speed is still a bottleneck for such systems. This effect is to some extent ameliorated by GPU utilization. Recent developments aim at reduction of detection times e.g. by cascading CNNs [5] or by detecting salient regions first using fuzzy logic [6] but significant reduction of training time is still an open area of research.

One contribution of the paper is the procedure of collecting as much realistic training data as possible providing limited user interaction. Ideally the system should be able to learn from a single ROI selection, all additional examples should be obtained automatically. Such least-user-effort approaches were already discussed e.g. for semi-automatic video annotation and detection systems, such as [7], [8]. In the cited approach, however, the user may be asked to annotated video several time (to decide about samples lying on decision boundary) which is not necessarily acceptable for all end users. An example of another successful detector that works on a single selection is given in [9]. The detector operates on sparse image representation (collection of SIFT descriptors) so it is very fast. Our initial experiments have shown that descriptor-based approaches works the best for highly textured and fairly complex objects.

The procedure of collecting training data given in this paper combines object tracking and background subtraction methods for semi-supervised collection of training windows together with foreground masks. The samples collected during tracking are further synthetically generalized (augmented) to enrich the training set. Scenarios, where tracking results are utilized for the collection of detector's training data, were already covered in literature, especially regarding tracking, with prominent examples [10], [11] or more recent CNN approaches [12], [13]. In such approaches the exact foreground-background separation (which is crucial for effective samples synthesis) is often neglected, since the algorithms typically have enough frames to collect rich training data.

The proposed methods were evaluated on a corpus of

TABLE I: Dictionary of abbreviations

Abbreviation	Expansion
CC	Cascade Classifier
CNN	Convolutional Neural Network
CSK	Circulant Structure of Kernels
EER	Equal Error Rate
FPR	False Positive Rate
GPU	Graphic Processing Unit
HD	High Definition
HOG	Histogram of Oriented Gradients
P-REACT	Petty cRiminality diminution through sEarch and Analysis in multi-source video Capturing and archiving plaTform
RBF	Radial Basis Function
RGB-D	Red Green Blue - Depth
ROC	Receiver Operator Characteristics
ROI	Region of Interest
SIFT	Scale Invariant Feature Transform
SURF	Speeded Up Robust Features
SVM	Support Vector Machine
TPR	True Positive Rate

surveillance videos and proved that its efficiency is good enough to be effective in supporting a user (police officer or security official) in their common working tasks.

The paper is organized as follows: in section II there is given a technical background and methodology used in our system, section III provides experimental results and IV contains conclusions. For reader's convenience Table I provides a short dictionary of abbreviations used in the paper.

## II. METHODS

### A. Detector overview

In the system described in this paper we utilize a classic detection framework, where a sliding window with varying sizes is moved over each frame and for each location the selected image part is evaluated against information gathered from training samples. A crucial part of the detector is formed by a SVM classifier which is responsible for evaluation of each selected image part. A pure SVM classifier when applied to hundred of thousands candidate areas would be too slow to learn and detect, so in our scenario so pre-classification step utilizing HAAR-like features-based cascade classifier is applied to limit the number of candidate windows to about several hundreds. We claim that this simple structure combines good detection rate together with acceptable detection speed (about 10 full-HD frames per second on modest Core i5 computer) as well as fine training speed in typical scenarios (up to few minutes).

In essence the two-stage detector architecture resembles some significant modern CNN approaches, where the detection is divided into region-proposal part and the region recognition part (see: e.g. [14]). In our approach region proposal is performed by cascade classifier, and final classification is done by SVM classifier. Both methods offer reasonable training and detection speeds required for this application.

In our scenario sources of data are naturally sparse. Depending on user decision the detector can be trained either on one or a short sequence of training images. Therefore a

critical part of our system are tools aiding user in an effortless collection of training examples from short image sequences as well as methods for artificial synthesis and generalization of training samples to provide the detector with the training data as rich as possible. These tools and methods are discussed in subsequent sections. The overall structure of the training procedure is given in Fig. 1.

### B. Collection of positive training samples

Although for some patterns (which include e.g. flat patterns) good detection results can be obtained using only one selected sample that is further generalized and synthesized into a set with larger variability, in most cases detection results highly depend on size and diversity of input training set. In the scenario discussed in this paper these properties of the training set can (at least partially) be achieved by collecting samples from a short sequence of input images. Our scenario is organized as follows: (1) a user select object of interest using rectangular area, (2) the application tracks the object in subsequent frames of the sequence (with optional manual reinitialization), (3) object foreground masks are established using motion information.

1) *Object tracking and foreground-background separation:* For tracking of rectangular area an optimized version of CSK tracker [15] that utilizes color-names features [16] is used.

As a result of the tracking procedure we obtain a sequence of rectangular areas that encompass the object of interest in subsequent frames. In most cases both object foreground as well as background will be present in the tracked rectangle. However, if the object is moving against moderately static background we can exploit motion information to effectively separate object foreground from background by background subtraction.

Let the tracking results be described by a sequence of rectangular areas  $\{R^1, \dots, R^T\}$  and let us denote coordinates of pixel  $i$  as  $p_i$ , color attributes for pixel  $i$  at time  $t$  as  $c_i^t$  and a mean of color attributes in the background as

$$\bar{c}_i = \frac{1}{n_i} \sum_{t: p_i \notin R^t} c_i^t \quad (1)$$

where averaging factor  $n_i$  is the number of frames where tracking window does not contain pixel  $i$  and can be computed as  $n_i = |\{t : p_i \notin R^t\}|$ .

Now we can specify a background training sequence for each pixel  $\{\hat{c}_i^t\}$

$$\hat{c}_i^t = \begin{cases} c_i^t & \text{if } p_i \notin R^t \\ \bar{c}_i & \text{if } p_i \in R^t \end{cases} \quad (2)$$

In accordance with the rule above, only pixels that at given time-step do not belong to the tracked area contribute to the background model computed for the image. Each pixel that always belong to tracked area is conservatively treated as foreground.

The background model adopted here follows algorithms from [17]. In this method scene color is represented independently for all pixels. The color for each pixel (both from

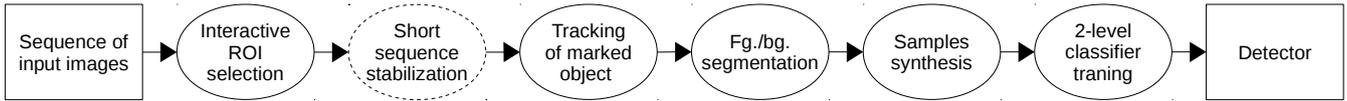


Fig. 1: Structure of the training procedure

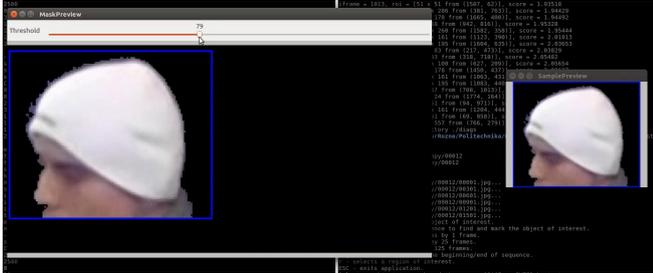


Fig. 2: Results of automatic foreground-background separation



Fig. 3: Division into a positive (P) and negative (N1-N4) examples

background and foreground  $BG + FG$ ) given the training sequence  $C_T$ , is modelled as:

$$p(c_i | C_T, BG + FG) = \sum_{m=1}^M \hat{\pi}_m \mathcal{N}(c_i; \hat{\mu}_m, \hat{\sigma}_m) \quad (3)$$

whereas the background model ( $BG$ ) is built from the selected number of largest clusters in the color mixture

$$p(c_i | C_T, BG) = \sum_{m=1}^B \hat{\pi}_m \mathcal{N}(c_i; \hat{\mu}_m, \hat{\sigma}_m) \quad (4)$$

where  $\hat{\mu}_m, \hat{\sigma}_m$  are estimated means and standard deviation of normal components in the mixture,  $\hat{\pi}_m$  are mixing coefficients  $M$  is the total number of mixtures and  $B$  is the selected number of foreground components. The pixel is decided to belong to the background when

$$p(c_i | BG) > c_{thr} \quad (5)$$

Threshold  $c_{thr}$  can be interactively adjusted by the user. Exact algorithms for updating mixture parameters are given in [17]. Sample result of background subtraction procedure is given in Fig. 2

2) *Image stabilization in a short sequence*: The foreground-background segmentation procedure works best when stable camera position is available or image sequence is stabilized before segmentation. The system proposed here uses a simple stabilization procedure basing on matching of SURF features [18] and computation of homography transformation between pairs of images. The stabilization works on short subsequences of the original sequence. First frame to stabilize is the frame used for marking the initial region of interest. The procedure then aligns all subsequent frames to the first frame by evaluating homographies relating two images. In order to do so, matching methods from [19] and the Least Median of Squares principle [20] is utilized. To increase stabilization efficiency GPU-accelerated procedures for keypoints/descriptors extraction and matching from OpenCV library are utilized [21].

### C. Collection of negative training samples

Negative samples that are used in detector training are extracted from the same sequence images that positive samples originated from. For each training image one fragment is used to extract positive sample, while the remaining part of the image is divided into at most four sources of negative samples as given in Fig. 3. Thus, an assumption is made that these remaining parts of the training sequence images do not contain positive samples. This assumption is not always valid, but may be strengthened by asking a user to mark **all** positive examples in the training sequence.

### D. Positive samples generalization and synthesis

1) *Geometric generalization*: In this step 3D rotations of patterns and their masks are applied to collected pattern images and their masks. It is assumed that patterns are planar, so this generalization method can be useful only to some extent for non-planar objects. The rotation effect is obtained by an applying a homography transformation, imitating application of three rotation matrices  $R_x(\alpha), R_y(\beta), R_z(\gamma)$  to a 3D object. The matrices correspond to rotations around  $x, y$  (in-plane rotations) and  $z$  (in-plane rotation) axes correspondingly. 3D rotation matrices are defined classically

$$\begin{aligned} R_x(\theta) &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{pmatrix} \\ R_y(\theta) &= \begin{pmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{pmatrix} \\ R_z(\theta) &= \begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{aligned} \quad (6)$$

To compute the transformation, first a homography matrix is computed using formula

$$H = R - \frac{\mathbf{t}\mathbf{n}^T}{d} \quad (7)$$

where  $\mathbf{n}$  is a vector normal to the pattern plane (we set it to  $\mathbf{n} = (0, 0, 1)^T$ ),  $d$  is the distance from the virtual camera to the pattern (we set it arbitrarily to  $d = 1$ , since it only scales 'real-world' units of measurement) and  $R$  is the 3D rotation matrix and can be decomposed as

$$R = (R_x(\alpha) \cdot R_y(\beta) \cdot R_z(\gamma))^{-1} \quad (8)$$

In order for the image center (having world coordinates  $C = (0, 0, d)^T$ ) to remain intact during transformation we define 'correcting' translation vector as

$$\mathbf{t} = -RC + C \quad (9)$$

Then we can specify artificial camera matrices as  $K_1$  and  $K_2$

$$K_1 = \begin{pmatrix} f & 0 & c_{in}^x \\ 0 & f & c_{in}^y \\ 0 & 0 & 1 \end{pmatrix}, K_2 = \begin{pmatrix} f & 0 & c_{out}^x \\ 0 & f & c_{out}^y \\ 0 & 0 & 1 \end{pmatrix} \quad (10)$$

where  $(c_{in}^x, c_{in}^y)^T$  and  $(c_{out}^x, c_{out}^y)^T$  are pixel coordinates of input and output image correspondingly, while  $f$  is the artificial camera focal length given in pixels. In this application we set  $f$  to be  $f_{mul}$  times larger input image dimension. Multiplier  $f_{mul}$  decides about the virtual distance of our virtual camera to the object. Smaller values introduce larger perspective distortions of the transformation, larger values introduce smaller distortions. We arbitrarily set  $f_{mul}$  to 10 implying only slight perspective distortions.

The final homography transformation applied to the pixels of the input image is given by

$$P = K_2 H K_1^{-1} \quad (11)$$

Rotation angles  $\alpha$ ,  $\beta$  and  $\gamma$  are selected randomly from the uniform distribution (denoted here as  $\mathcal{U}$ ). The amount of rotation around axes  $y$  is twice times the amount of rotation around remaining axes to better reflect dominant rotations in human movement

$$\begin{aligned} \alpha &\sim \mathcal{U}(-1, 1) \cdot \delta_{max} \cdot 0.5, \\ \beta &\sim \mathcal{U}(-1, 1) \cdot \delta_{max}, \\ \gamma &\sim \mathcal{U}(-1, 1) \cdot \delta_{max} \cdot 0.5 \end{aligned}$$

and  $\delta_{max}$  is the parameters specifying the maximum extent of allowed rotation.

2) *Intensity and contrast synthesis*: In the proposed approach image intensity and contrast synthesis is applied in addition to geometric transformations. It is especially important for Haar-like features that lack intensity normalization. A simple linear formula is used here. For each pixel gray value  $I_{in}$  we have

$$I_{out} = a * I_{in} + b \quad (12)$$

where

$$a = 1 + c_{dev}, b = I_{dev} - \mu_I \cdot c_{dev} \quad (13)$$

where  $\mu_I$  is the average intensity of the sample and contrast deviation  $c_{dev}$  as well as intensity deviation  $I_{dev}$  are sampled from the uniform distribution  $c_{dev} \sim \mathcal{U}(-1, 1) \cdot c_{max}$  and  $I_{dev} \sim \mathcal{U}(-1, 1) \cdot I_{max}$ .  $c_{max}$  is a parameter denoting the maximum allowed contrast change and  $I_{max}$  is a parameter denoting the maximum allowed intensity change. Changes in contrast preserve mean intensity of an image. After application of the formula its results are appropriately saturated.

3) *Application of blur*: Training and test samples may differ in terms of quality of image details due to different factors such deficiencies of optics used or motion blur. In our case we apply a simple Gaussian filter in order to simulate natural blur effects

$$\sigma = \mathcal{U}(0, 1) \cdot \sigma_{max} \cdot \min(I_{width}, I_{height}) \quad (14)$$

where  $I_{width}$  and  $I_{height}$  are image sample sizes and  $\sigma_{max}$  controls the maximum size of the Gaussian kernel.

4) *Merging with the background*: Generalized training images are superimposed on background samples extracted from negative examples of size ranging from about 0.25 to 4 times the positive sample size. Gray-level masks are used for seamless incorporation of positive samples into background images.

### E. Detector training

Before training all training samples are resampled to a fixed size of 24x24 pixels. The detector training procedure is divided into two steps. In the first step the cascade classifier using HAAR-like features is trained. In our scenario for each cascade stage 300 positive samples and 100 negative samples are utilized. Minimum true positive rate for each cascade level is set to 0.995 and maximum false positive rate is set to 0.5. The classifier is trained for a maximum of 15 stages or until reaching  $\approx 0.00003$  FPR. The expected TPR is at least  $0.995^{15} \approx 0.93$ . By using these settings up to about 1000 detections are generated for each Full-HD test image.

During the second stage of training an SVM classifier is trained to handle samples that passed the first cascade classification. For most experiments the SVM classifier is trained on 300 positive and 300 negative samples. The SVM classifier uses Gaussian RBF kernel.

$$K(\mathbf{x}, \mathbf{y}) = \exp(-\gamma \|\mathbf{x} - \mathbf{y}\|^2) \quad (15)$$

The Gaussian kernel size  $\gamma$  and SVM regularization parameter  $C$  are adjusted using automatic cross-validation procedure performed on the training data. For SVM classification Histogram of Oriented Gradients features [22] are extracted. For each sample a 9-element histogram in 4x4 cells is created with 16x16 histogram normalization window overlapping by 8 pixels, thus giving  $4 * 16 * 9 = 576$  HOG features in total.

Negative samples are extracted from Cascade Classifier decision boundary (containing samples that were positively verified by CC but still negative) if possible. If not - image

fragments used as background images for positive samples or other randomly selected samples are used. In all experiments OpenCV 3.1 [21] Cascade Classifier and SVM implementation are utilized.

Given our test data, the number of resulting support vectors in the SVM classifier varies between 200 and 400. Let us review one specific configuration: 'hat' pattern trained on 55 images with masks and pattern generalization settings  $\sigma_{max} = c_{max} = 0$ ,  $\delta_{max} = 0.7$ ,  $I_{max} = 50$ . After SVM metaparameter optimization we obtain SVM regularization parameter  $C = 2.5$ , RBF kernel size  $\gamma = 0.5$  and the number of support vectors 233.

#### F. Detection and post-processing

During detection phase each test image is first processed by the cascade classifier typically returning several hundreds candidate areas. After this, each candidate area is examined by the SVM classifier and a score is assigned to each detection. The score is computed as the signed distance from the separating plane in support vector space with lowest negative scores treated as best matches and high positive scores as worst matches.

For each image only the best score area is considered for further processing. Frames from the test sequence are sampled and processed with increasing density (first, last and middle frame for start and then intermittent frames), to quickly produce some results for the user to review (non minima suppression is used to reduce clutter)

### III. EXPERIMENTS

#### A. Preliminary experiments

During the first stage of experiments there was selected a single test sequence '00012' with 1776 Full-HD frames. Using this sequence various parameter configurations were evaluated in order to assess basic properties of the solution proposed. Basing on these experiments some answers can be given regarding problems such as impact of utilization of two-layer detector on detection results and detection/training speed, impact of the method of selection of training samples on detection accuracy or impact of values of image synthesis parameters on overall quality. Above questions will be discussed in the following paragraphs. All experiments were performed on Intel Core i5 computer. During the first 3 experiments one sample pattern 'hat' was utilized, in the last experiment 3 other patterns 'logo', 'helmet' and 'shirt' were introduced. Examples of training samples are given in Fig. 4 and samples marked in full-frame image are given in Fig. 5. Filtered detection results for one test sequence presented in the form of a simple GUI are given in Fig. 6.

a) *Two-layer detector*: In the first experiment there was evaluated a trade-off between detection and training speed for different number of expected cascade stages  $k$  (Fig. 8). Identical parameters were used for all  $k$  except for the number of SVM training samples. For  $k < 15$  there were used 900 positive and negative samples to accommodate for weaker selectivity of the 1-st detection stage. For  $k \geq 15$  the default of

300 positive and negative samples were utilized as in all other experiments. The experiment shows that for low  $k$  training time is dominated by SVM training, for large  $k$  cascade training dominates. A good compromise for our data can be obtained for  $k = 15$ . Larger  $k$  obviously means also faster detection (Fig. 7), but also slightly worse detection results (Fig. 9) (likely due to utilization of more robust HOG features in the second stage).

b) *Collection of training samples*: In the next experiments there were compared detector performance for different training data collection methods. In the first place the data samples were collected using automatic tracking and foreground-background separation methods given in this paper. In the process 55 data samples from of 'hat' pattern were collected together with their automatically generated masks. The data consisted of images of a hat on top of a head, while the head was making full 180 degrees rotation around central axis. For comparison, a short sequence of training samples representing only 3 extreme head positions (*en-face* and two profiles) was utilized. For both sequences either appropriate foreground-background masks or no masks were used giving 4 different combinations of settings. The detection results are given in Fig. 10.

Not surprisingly the richest possible data source (55 frames with generated masks) gives the best results. It is valuable to note that for our data, application of both object tracking and automatic mask generation is substantial to get optimal results.

c) *Synthetic generalization of training data*: In these experiments different measures and intensities of samples synthesis were evaluated. The results are given in Fig. 11 and Fig. 12. The results show that moderate geometric as well as contrast and sharpness generalization provides best results. However, the selection of appropriate parameters is object and sequence-specific. E.g. it may be observed that near-flat surfaces e.g. 'logo' benefits from aggressive geometric distortions (i.e. larger rotation angles). In addition, the reduction of sharpness proved to work best for computer-graphics-generated samples.

d) *Detection of various patterns*: In the last of our preliminary experiments there was evaluated how the detector handles different types of patterns. Therefore, the pattern 'logo' was trained on a single training example with no mask, the pattern 'shirt' was trained on a sequence of 30 samples without a mask and the pattern 'helmet' was trained on 41 samples also without a mask. The result are given in Fig. 13.

It can be noted the relatively worse performance for the 'shirt' pattern, mainly due to numerous occlusions. Even in the case of the 'shirt' pattern we still have about 90% of successful hits for recall rates of 0.3. For best patterns such as 'helmet' we have 70% of positive examples with still 0 false positives!

In the course of experiments, it was observed that motion blur (inherent or originating from de-interlacing) is the most destructive type of noise regarding both training and detection phase. In addition, due to quite severe subsampling of the pattern (down to  $24 \times 24$ ), the detector may suffer from problems in distinguishing between patterns differing only in



Fig. 4: Example training samples of 'hat', 'logo', 'helmet' and 'shirt'



Fig. 5: Frame with marked 'hat', 'logo', 'helmet' and 'shirt' samples

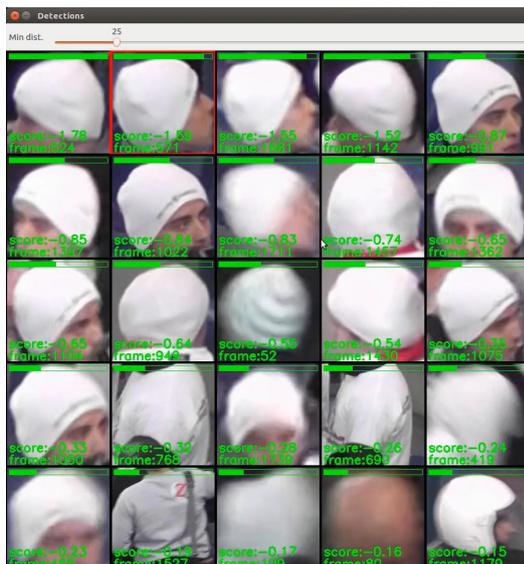


Fig. 6: Detection results filtered by minimum distance (25 frames) between hits

small details. On the other hand, due to this property, the detector should well handle also small patterns - only slightly bigger than the nominal  $24 \times 24$  pattern size.

### B. Large-scale experiments

Tests of the presented algorithm were conducted on a dataset containing 11 recordings, with nearly 30 thousand frames in total, with full HD resolution. Three patterns were created

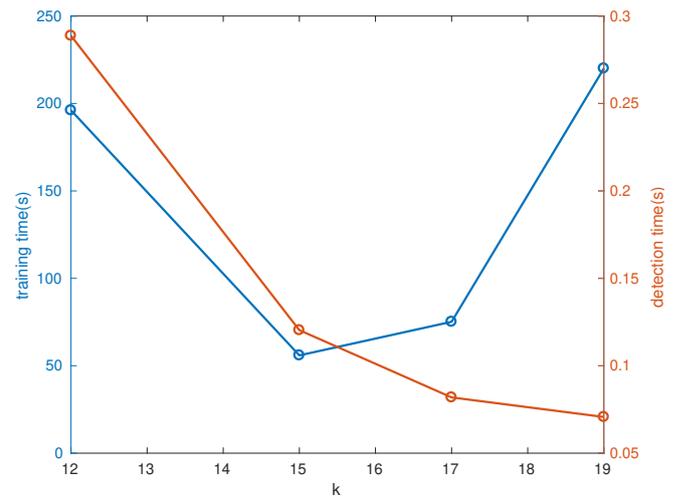


Fig. 7: Training/detection time vs. the number of cascades ( $k$ ).

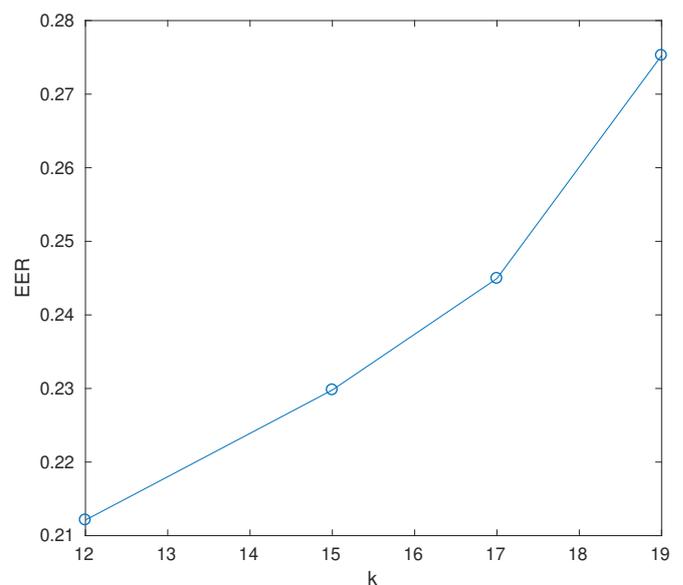


Fig. 8: 'hat' detection EER vs. the number of cascades ( $k$ )

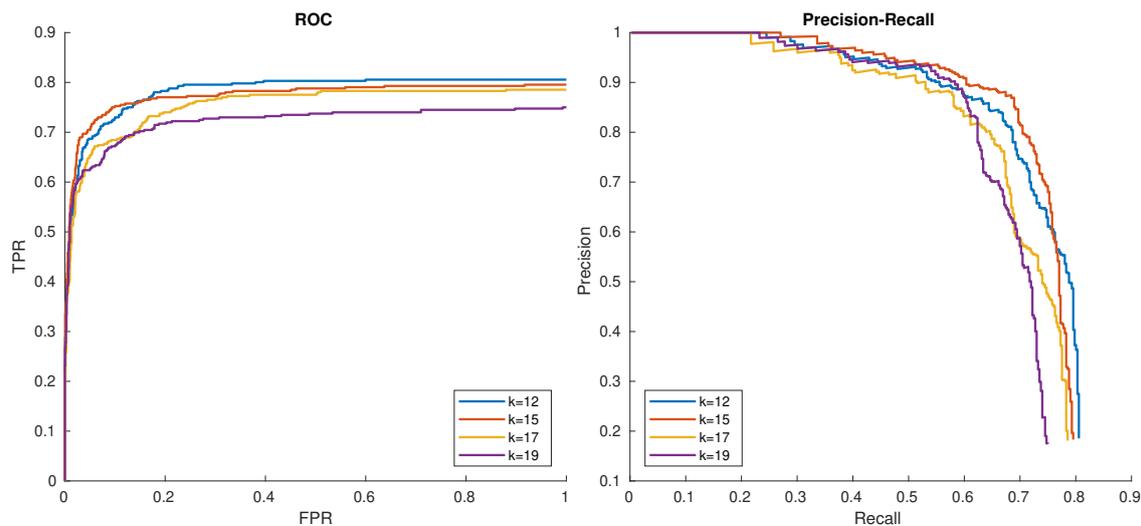


Fig. 9: 'hat' in '00012' detection results with respect to number of the requested cascade stages.

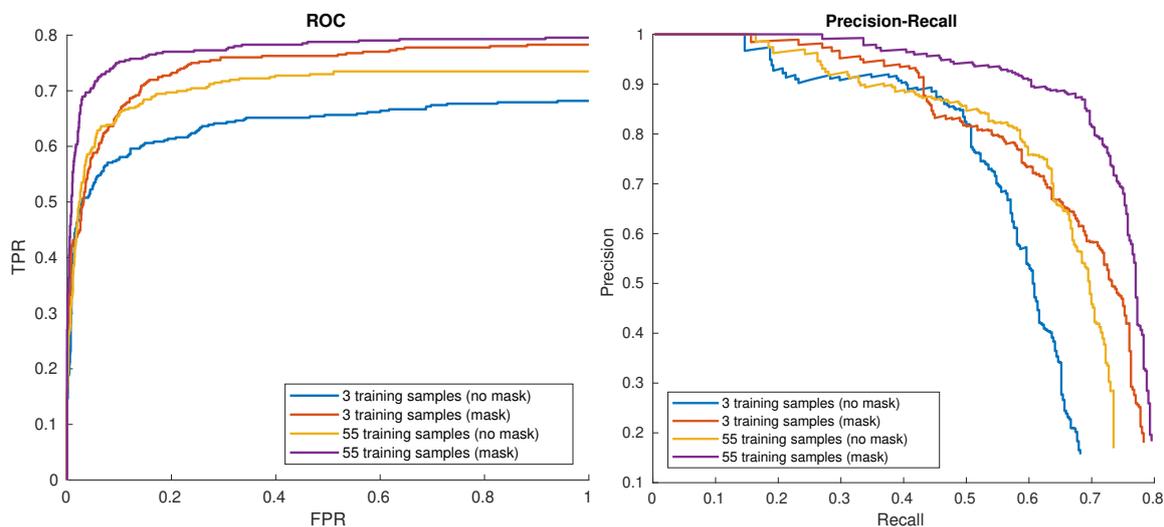


Fig. 10: 'hat' in '00012' detection results for different training data collection methods

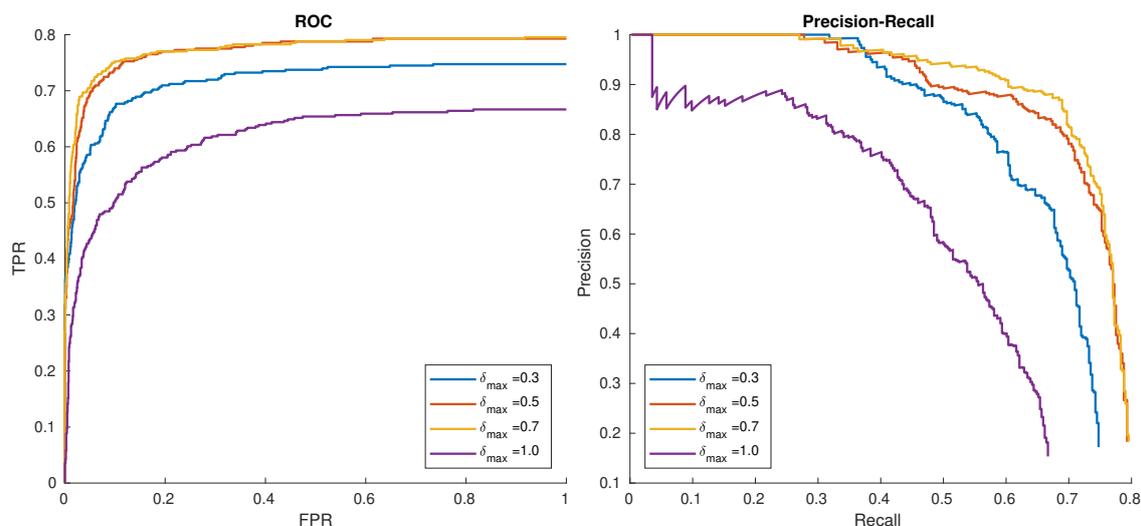


Fig. 11: 'hat' in '00012' detection results for different levels of geometric synthesis

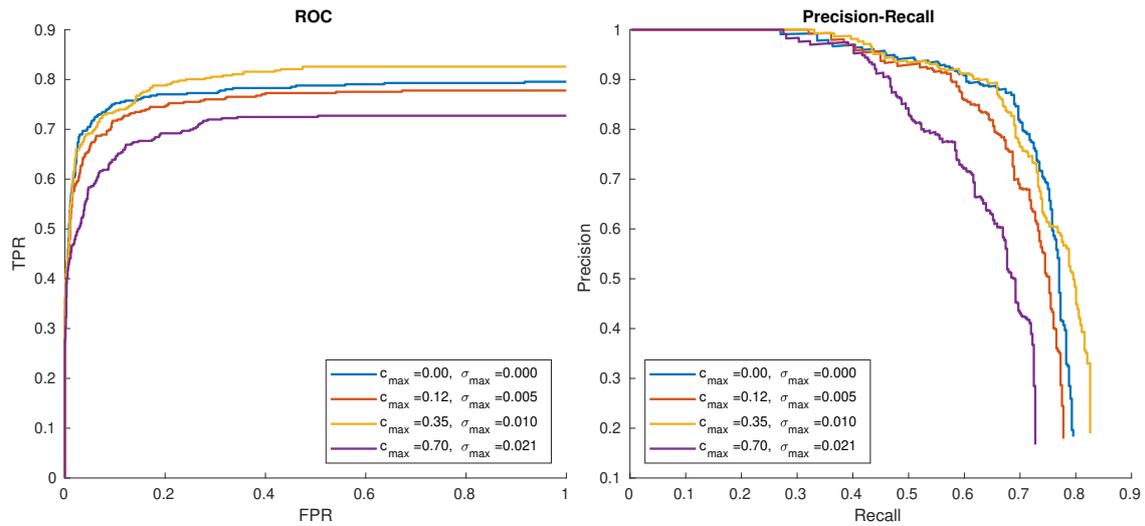


Fig. 12: 'hat' in '00012' detection results for different contrast and sharpness synthesis levels

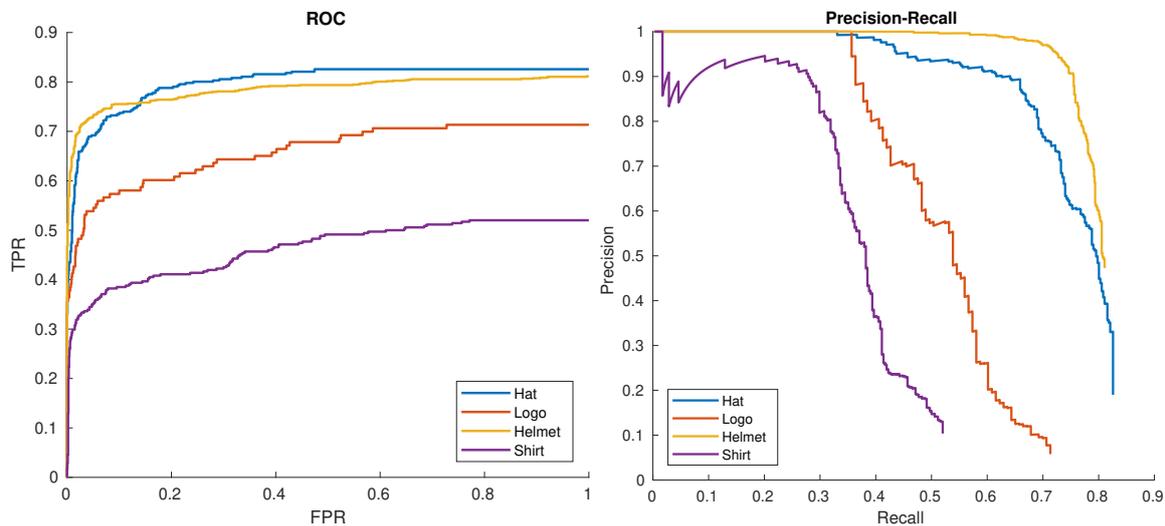


Fig. 13: ROC and PR curves of 'hat', 'logo', 'helmet' and 'shirt' detections in '00012' sequence

(Fig. 14), and all sequences were carefully labeled by hand to create ground-truth data. All patterns were created based on a single frame (one positive sample). As a training data, high quality still picture was used, with resolution scaled down to full HD.

Results of the experiments (ROC curve) for the selected pattern A is presented on Fig. 15a. EER is similar for all patterns A,B,C, and is equal to 25.3%, 28.3% and 28.0% for each pattern respectively. Accumulated EER equals to 27.4%. Obtained results resemble those from small dataset. Even though the training sample and query images were taken with different devices and had different quality, the algorithm gave satisfactory results.

Final addition to the testing scenario was the utilization of short sequences. For every short sequence, from all the results only the one with the best response was taken as a final detection and passed to further processing. Accumulated

results for the sequences of length 5 is presented on Fig. 15b (remaining charts are given in supplemental materials). EER for them are, respectively: 27.4%, 15.1% and 14.3%. It was observed that the longer the sequence the smaller is the quality gain.

More tests were also conducted using one of the widely used dataset – RGB-D Object Dataset [23]. It contains multiple everyday objects, along with masks, that can be used to create models and short sequences of scenes with multiple objects. Fig. 15c presents sample results obtained for the cereal\_1 object in desk\_3 sequence. Model was created using only 7 views of the object in this case.

#### IV. CONCLUSIONS

In this paper, we presented a solution that can support work of police officers in surveillance tasks. The system proved to positively address difficult task requirements concerning sparse



Fig. 14: Selected test patterns

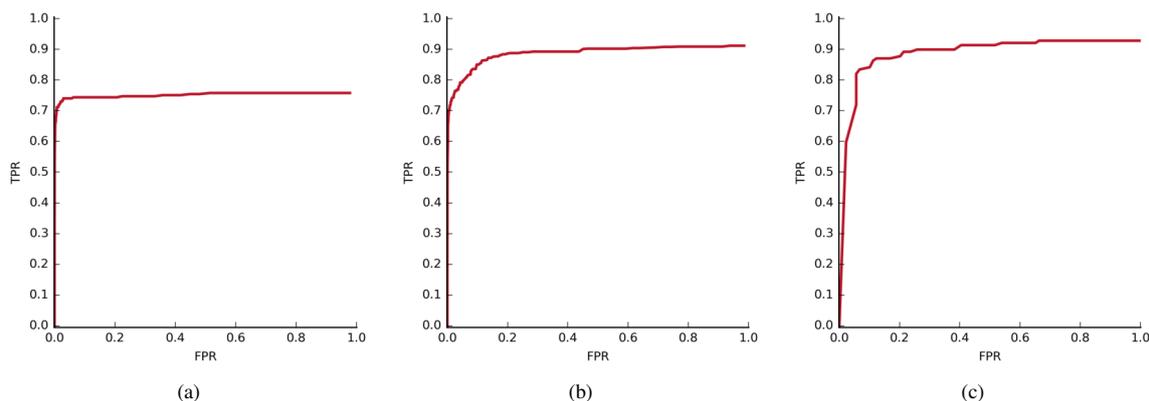


Fig. 15: (a) ROC curve for the pattern A. (b) Accumulated ROC curve for 5-elements sequence analysis. (c) ROC curve for cereal\_1 object in desk\_3 sequence

training data, quick learning and fast and reliable detection. An attractive training/detection speed and recognition rate trade-off was obtained by the application of 2-layer cascade/SVM classifier. The system proposed can learn from a single training sample, but also can collect samples from short image sequences with only small user supervision in order to obtain rich training data. Performance of the system vary depending on the type and quality of training/test data, but we argue that on average results are satisfactory and even not-the-best results provide sufficient information to be useful in practical surveillance scenario.

#### REFERENCES

- [1] J. Arraiza, N. Aginako, G. Kioumourtzis, G. Leventakis, G. Stavropoulos, D. Tzovaras, N. Zotos, A. Sideris, E. Charalambous, and N. Kouras, "Fighting volume crime: an intelligent, scalable, and low cost approach," *9th Summer Safety & Reliability Seminars, SSARS 2015, June 21- 27, 2015, Gdansk/Sopot, Poland*, 2015.
- [2] S. Blunsden and R. Fisher, "The behave video dataset: Ground truthed video for multi-person behavior classification," *Annals of the BMVA*, vol. 2010, no. 4, pp. 1–11, 2010.
- [3] G. Awad, C. G. M. Snoek, A. F. Smeaton, and G. Quénot, "Trecvid semantic indexing of video: A 6-year retrospective," *ITE Transactions on Media Technology and Applications*, vol. 4, no. 3, pp. 187–208, 2016. doi: 10.3169/mta.4.187 Invited paper.
- [4] J. Redmon, S. K. Divvala, R. B. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *CoRR*, vol. abs/1506.02640, 2015. [Online]. Available: <http://arxiv.org/abs/1506.02640>
- [5] D. Zeng, F. Zhao, S. Ge, and W. Shen, "Fast cascade face detection with pyramid network," *Pattern Recognition Letters*, vol. 119, pp. 180 – 186, 2019. doi: <https://doi.org/10.1016/j.patrec.2018.05.024> Deep Learning for Pattern Recognition. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0167865518302125>
- [6] M. Woźniak and D. Połap, "Object detection and recognition via clustered features," *Neurocomputing*, vol. 320, pp. 76 – 84, 2018. doi: <https://doi.org/10.1016/j.neucom.2018.09.003>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0925231218310634>
- [7] Y. Abramson and Y. Freund, "Active learning for visual object detection," UCSD, Tech. Rep., 01 2006.
- [8] —, "SEmi-automatic Visual LEarning (SEVILLE): Tutorial on active learning for visual object recognition," *Proc. CVPR*, 2005.
- [9] J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," in *Proceedings of the Ninth IEEE International Conference on Computer Vision - Volume 2*, ser. ICCV '03. Washington, DC, USA: IEEE Computer Society, 2003. ISBN 0-7695-1950-4 pp. 1470–. [Online]. Available: <http://dl.acm.org/citation.cfm?id=946247.946751>
- [10] Z. Kalal, K. Mikołajczyk, and J. Matas, "Tracking-learning-detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1409–1422, July 2012. doi: 10.1109/TPAMI.2011.239
- [11] M. Andriluka, S. Roth, and B. Schiele, "People-tracking-by-detection and people-detection-by-tracking," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, June 2008. doi: 10.1109/CVPR.2008.4587583. ISSN 1063-6919 pp. 1–8.
- [12] C. Feichtenhofer, A. Pinz, and A. Zisserman, "Detect to track and track to detect," in *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*, 2017. doi: 10.1109/ICCV.2017.330 pp. 3057–3065. [Online]. Available: <https://doi.org/10.1109/ICCV.2017.330>
- [13] K. Kang, W. Ouyang, H. Li, and X. Wang, "Object detection from video tubelets with convolutional neural networks," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 817–825, 2016.

- [14] S. Ren, K. He, R. B. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *CoRR*, vol. abs/1506.01497, 2015. [Online]. Available: <http://arxiv.org/abs/1506.01497>
- [15] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "Exploiting the circulant structure of tracking-by-detection with kernels," in *Computer Vision – ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part IV*, A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. doi: 10.1007/978-3-642-33765-9\_50. ISBN 978-3-642-33765-9 pp. 702–715. [Online]. Available: [https://doi.org/10.1007/978-3-642-33765-9\\_50](https://doi.org/10.1007/978-3-642-33765-9_50)
- [16] M. Danelljan, F. S. Khan, M. Felsberg, and J. v. d. Weijer, "Adaptive color attributes for real-time visual tracking," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, June 2014. doi: 10.1109/CVPR.2014.143. ISSN 1063-6919 pp. 1090–1097.
- [17] Z. Zivkovic, "Improved adaptive gaussian mixture model for background subtraction," in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, vol. 2, Aug 2004. doi: 10.1109/ICPR.2004.1333992. ISSN 1051-4651 pp. 28–31 Vol.2.
- [18] H. Bay, T. Tuytelaars, and L. Van Gool, *SURF: Speeded Up Robust Features*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 404–417. ISBN 978-3-540-33833-8. [Online]. Available: [https://doi.org/10.1007/11744023\\_32](https://doi.org/10.1007/11744023_32)
- [19] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov 2004. doi: 10.1023/B:VISI.0000029664.99615.94. [Online]. Available: <https://doi.org/10.1023/B:VISI.0000029664.99615.94>
- [20] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection*. John Wiley & Sons, Inc., 2005, ch. Algorithms, pp. 197–215. ISBN 9780471725381. [Online]. Available: <http://dx.doi.org/10.1002/0471725382.ch5>
- [21] Itseez, "Open source computer vision library," <https://github.com/itseez/opencv>, 2015.
- [22] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, June 2005. doi: 10.1109/CVPR.2005.177. ISSN 1063-6919 pp. 886–893 vol. 1.
- [23] K. Lai, L. Bo, X. Ren, and D. Fox, "A large-scale hierarchical multi-view rgb-d object dataset," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1817–1824.

# Robust Image Forgery Detection Using Point Feature Analysis

Youssef William   
German University in Cairo  
Cairo, Egypt  
youssef.teryak@student.guc.edu.eg

Sherine Safwat   
German University in Cairo  
Cairo, Egypt  
sherine.safwat@guc.edu.eg

Mohammed A.-M. Salem   
German University in Cairo,  
Ain Shams Univeristy  
mohammed.salem@guc.edu.eg,  
salem@cis.asu.edu.eg  
Cairo, Egypt

**Abstract**—Day for day it becomes easier to temper digital images. Thus, people are in need of various forgery image detection. In this paper, we present forgery image detection techniques for two of the most common image tampering techniques; copy-move and splicing. We use match points technique after feature extraction process using SIFT and SURF. For splicing detection, we extracted the edges of the integral images of  $Y$ ,  $C_b$ , and  $C_r$  image components. GLCM is applied for each edge integral image and the feature vector is formed. The feature vector is then fed to a SVM classifier. For the copy-move, the results show that SURF feature extraction can be more efficient than SIFT, where we achieved 80% accuracy of detecting tempered images. On the other hand, processing the image in  $YC_bC_r$  color model is found to give promising results in splicing image detection. We have achieved 99% true positive rate for detecting splicing images.

**Index Terms**—Image Forgery, Copy-Move Forgery Detection, Image Splicing, SIFT, SURF, Support Vector Machine (SVM), benchmark dataset, CASIA datasets, Gray Level Co-occurrence Matrix (GLCM)

## I. INTRODUCTION

IN today's world, digital images are widely used in various domains such as; newspapers, scientific journals, magazines, and many other fields [25]. Unfortunately, today's digital technology made it easy for digital images to be forged due to the availability of the low cost photo editing software [17]. For example, during the incident of Hurricane Harvey, fake images were posted of sharks inside New York as shown in Figure 1. Another example, in Figure 2 as the cutout of the newspaper showed a forged photographs of Bill Clinton, and Saddam Hussein at the White House [9].

Of course, this can cause chaos and panic among the viewers of such digital images. In addition, it can cause erosion in people's trust towards images [20]. Thus, in order to recover people's trust towards digital images, it is important to develop new trustworthy techniques for digital images forgery detection.

Image forgery detection is such a complicated job. Nowadays, it became very difficult to detect whether an image is fake or not. According to Huynh, et al. image forgery detection is one type of the passive techniques that use blind algorithms

for tampering detection in the suspected image without using any prior information. Accordingly, they divided passive techniques into two types: copy-move and splicing [11].



Figure 1. Hurricane Harvey fake reports that were published in BBC in 2017

The copy-move is defined by copying region of an image and pasting it in another place in the *same* image, generally to hide unwanted parts of the image. On the other hand, image splicing is the process of copying a region of an image and pasting it in another place in *another* image. Thus, detection of tampered regions is done through searching for very similar regions in copy-move images and completely odd regions in spliced images [12].

In this paper, we are extracting the image features and analyzing it to detect the forged images and also determine the type of the forgery whether it is copy-move or splicing.



Figure 2. Example of realistic looking forgery

Our work is test on multiple datasets. The rest if the paper is organized as follows; in Section II, we present the literature review for the copy-move and image splicing forgery detection. Section III, introduces our Methodology for both copy-move and splicing along with the datasets used. Experimental results for both techniques are elaborated in Section IV. In Section V we discuss the results and the limitation of the proposed algorithms. Finally, in Section VI we drive the conclusions.

## II. LITERATURE REVIEW

This section summarizes some of the work done in copy-move and splicing detection as follows.

### A. Copy-Move Forgery Detection

The copy-move attack is one type of tampering in which a region of the image is copied and pasted in another area in the same image to cover an important image feature. In [25] a technique for detecting copy-move forgery is presented based on SURF and KD-Tree for multidimensional data matching. Shivakumar et al. designed a system to identify the duplicated areas, then extracted key points in the forged areas and matched them among the SURF features, thus determined the possibility of forgery.

Alberry et al. introduced a fast technique optimizing SIFT and fuzzy c-means clustering for copy-move forgery detection. First, the algorithm detected and matched the key points in the image and clustered the points based on their descriptors using c-means algorithm. Their algorithm could successfully come over the computational complexity in the matching stage after using the clustering algorithm [5].

Pasquini et al. designed an empirical system to verify online news by analyzing images from news article. The system identified the set of meta-data visuals related to the same topic and presented some common visual elements. After that, the data set was compared with many websites with the same topic. Thus, the system the could differentiate between the images and output the fake one [20].

In [9] Fridrich et al. succeeded in detecting the forged parts even when the copied areas were skillfully enhanced and merged with the background and saved in the lossy JPEG format. They introduces a novel correlation between the original image segment and the pasted part to be used as a basis for a successful detection for the copy-move.

The paper [7] examined several block-based methods to detect the copy-move forgery. Bayram et al. showed their time complexity and robustness in the results. They discussed Discrete Cosine Transform (DCT), Fourier Mellin Transform (FMT) and Principal Component Analysis (PCA). The results were good on any JPEG image, but the algorithm is limited to non-rotated or scaled objects. However, they could improve the efficiency of copy-move forgery techniques by counting

bloom filters, especially when the image quality is high.

Ryu et al. [22] proposed a forensic technique to localize duplicated image regions based on Zernike moments of small image blocks. They utilized the characteristics of rotation in variance to reliably unveil duplicated areas after random rotations. By examining the image, they designed a new block matching operation centered on locality-sensitive hashing and decrease fake positives. Their experiments indicated high robustness for JPEG compression, blurring, additive white Gaussian noise, and moderate scaling.

The work done [15] by Kakar et al. proposed a novel technique based on transform-invariant features for copy-move detection. The results provided efficacy of this technique in detecting copy-move forgeries with translation, scaling, rotation, flipping, lossy compression, noise addition and blurring.

Lin et al. [18] introduced an image forgery detection using both copy-move and splicing forgeries detector. They first used a forgery picture identification strategy through periodicity assessment with the double mixing impact in the temporal and DCT domain. Then the function obtained by SURF descriptors is implemented to resist the variety of rotating and/or scaling of tampered objects in an image. Experimental results showed that their suggested methods were well conducted in the identification of forgery location. The suggested methods were prepared to identify the forged areas and acknowledge the non-original areas, especially for the copy-move forgery pictures.

Finally, We built our work of copy-move detection on [8]. Christlein et al. examined the 15 most prominent feature sets and created a challenging real-world copy-move dataset "Benchmark", that we used as part of our dataset. The paper showed many algorithms in detecting copy-move forgery using both key-point and block-based methods. The results showed that key-point methods have a clear advantage in terms of computational complexity, while the most accurate detection was achieved through the block-based method Zernike.

### B. Image Splicing Forgery Detection

The splicing attack is one type of tampering in which different regions of the same or separate sources are combined to create a new fake image. In [21], Riess et al. introduced a method for detecting image splicing through the change of illumination environment of the spliced object. They could overcome one of the biggest challenges which is computing the lighting environment from homogeneous materials. Their approach could successfully improve the mean error by almost 30%. Yet, hair, structurally unsmooth regions, and highly textured clothes were from the model limitations.

Ke et al. proposed forged image detection technique based on shadow consistency, assuming that the shadow and the main body were copied from one image and pasted to another. The algorithm worked as follows; the suspicious region including shadow and non-shadow were first selected and the texture features were then extracted. Next, the similarity of the two texture characteristics were measured using the correlation function. Finally, by comparing the similarity, the decision would be made whether the image was tampered or not [16].

Similarly, an algorithm for digital image forgery detection based on shadow detection of the spliced object was presented in [26] by Tuba et al. They based their algorithm on the fact that a shadow wouldn't change the surface texture, thus if two adjacent areas (with and without shadow) had different texture, then the image could very likely be forged. The algorithm used Local Binary Pattern (LBP) from shadow areas and adjacent non-shadow areas. The energy and entropy extracted from the features histograms proved to be the most discriminating.

On the other hand, Hakimi et al. used different approach for detecting image splicing based on LBP and Discrete Wavelet Transform (DWT). The images were first converted from RGB into  $YCbCr$  color channel. Next, the chrominance component were divided into non-overlapping blocks. After that, LBP operator was performed and the wavelet transform applied to all blocks. The output was then fed to the Support Vector Machine (SVM) classifier as features. Haar wavelet was used to reduce the image dimension. The results showed that the algorithm was effective in detecting spliced photos with acceptable accuracy [10].

Regarding LBP, and DCT, Bebis et al. [4] proposed a method to detect image splicing forgeries using these two techniques. They divided the chrominance component of the input image into overlapping blocks, then once used 2D DCT and once used the LBP for each block. Standard deviation is then estimated along with the DCT or LBP to extract the feature vectors from each block and fed it to SVM. Their experiments were on Benchmark dataset with detection accuracy of 97%.

In [13] Huynh-Kha et al. focused on developing a system to detect copy-move and the splicing forgeries together in one image. By applying one-level Discrete Wavelet Transform, the sharpened edges with high frequencies were detected from LH, HL and HH sub-bands. The suspicious region was extracted the feature using Run Difference Method (RDM).

Wang et al. [28] worked on splicing detection through using the GLCM and detecting edges from the integral image and then passing the resulted features to a SVM classifier. They used all images component  $YCbCr$  in extracting the feature vectors of an image. They used a certain algorithm to detect

the edges of the image horizontally, vertically and diagonally. We built our work in splicing detection on this paper, we used integral image in detecting the edges of the image.

### III. METHODOLOGY

This section is divided into two subsections; copy-move detection technique, and the splicing detection technique. We will explain how our algorithm in both techniques, the workflow, and our datasets are represented in the block diagram, Figure 3.

#### A. Proposed Method of Copy-Move Forgery Detection

1) *Working Plan*: In copy-move detection, based on [8]. Given an image, the detected regions are computed through the following steps:

- Step 1: Convert the image from RGB to gray-scale color model.
- Step 2: Divide the image into 4 equal blocks and calculate their integral features.
- Step 3: Divide each of the 4 blocks into another four blocks of same size and execute their features.
- Step 4: Extract key-points of all blocks using SIFT and SURF.
- Step 5: Calculate a feature vector for each key-point.
- Step 6: Match each feature vector by comparing each block's features executed with another block.
- Step 7: The forgery is then detected according to a certain threshold among all blocks.
- Step 8: The detected blocks are then displayed with the common object plotted.

2) *Datasets of Copy-Move Images*: We used multiple datasets for copy-move detection; **MICC-F8multi** consisting of 8 forged PNG images, **MICC-F220** consisting of 220 images, 210 original images and 10 fake images [14]. Images were either scaled or rotated or duplicated in different parts of the image. The last dataset was the **Benchmark** datasets that consisted of 4 datasets [8]. Examples of Benchmark datasets are shown in Figure 9.

3) *Pre-processing*: In the beginning, our system was designed using MATLAB, where it requests an RGB image of any format, then the system converts it into a gray-scale. Then the image now is ready for the blocking process. A simple two stages algorithm is then used to divide an image into blocks. In the first stage, the image is divided to 4 equal blocks of the same size and angle. Similarly in the second stage, the system divides each individual block into another 4 equal smaller blocks. This approach is called "*Multi Staged blocking*". We will result in having 20 blocks (4 large blocks + 4\*4 small blocks) as shown in Figure 6. The blocking technique eases the features extraction and matching processes that will be discussed later.

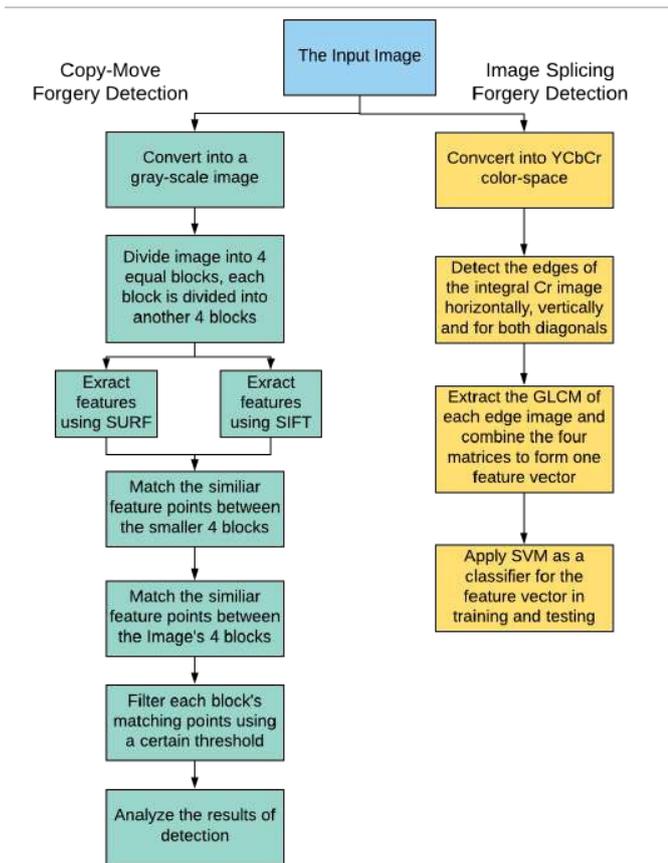


Figure 3. Image Detection Block Diagram

4) *Features Extraction*: For the features extraction, both key-point based methods were used; SIFT & SURF approaches for each block.

**SIFT Key-points based method**: SIFT (Scale-Invariant Feature Transform) is an algorithm to detect and describe local features in an image. The SIFT algorithm converts an image into a local feature vector called SIFT descriptors and these descriptors have powerful geometric transformations that are constant to scaling and rotation [5], [19].

In addition to extracting the features using SIFT, Harris features on the gray image is used to find the corner points. This process is applied to each block of the image. As a result, we obtain the valid points for the neighboring features.

**SURF Key-points based method**: similar to SIFT, SURF (Speed Up Robust Feature) is a descriptor used to recognize and locate objects. The values of Hessian determination for each pixel in the image are used to find the points of interest. Next, functions are constructed to be used to select extreme points [6].

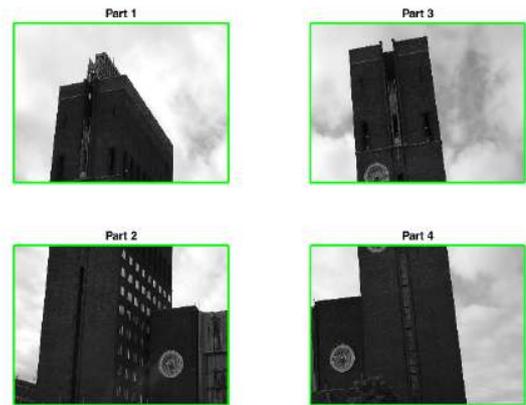


Figure 4. represents the first stage in multi-blocking

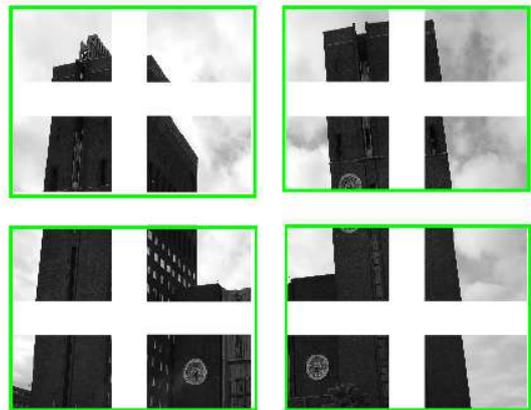


Figure 5. represents the second stage in multi-blocking

Figure 6. An example for the multi-stage blocking of a gray-scale image.

Alternatively, we replace the SIFT step with the SURF. Then, we find the corner points using the Harris detection on the gray image. This process is performed on each block of the image. Lastly, we obtain the valid points for the neighboring features.

5) *Matching Points*: After extracting the neighboring features of each block, the neighboring features are compared to features of another blocks as to find the matched features. Successfully, the locations of the corresponding points for each block will be determined. Ultimately, the system allows the user to view the corresponding points. The system shows the two suspicious blocks where they exceeded the threshold of detected matched points as shown in Figure 7.

6) *Filtering & Analyzing*: The blocks are filtered according to a threshold for the number of matching points detected between two blocks. The threshold is calculated from the

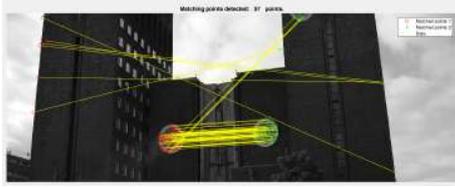


Figure 7. Some of the detected blocks of the image

average number of matched points detected in our datasets.

The system calculates a percentage of the forgery in the image based on the number of suspicious blocks. Accordingly, the percentage of forgery decides which key-point-based method works better on our datasets.

### B. Proposed Method of Image Splicing Forgery Detection

Regarding the image splicing forgery detection our algorithm is based on the Gray Level Co-occurrence Matrix (GLCM) for feature extraction similar to [28] and the Support Vector Machine (SVM) for classification [23].

1) *Working Plan*: Given an RGB image as an input, our system runs as follows:

- Step 1: Convert the RGB image to the  $YC_bC_r$  image component.
- Step 2: Extract each color channel.
- Step 3: Edge detection is performed on each individual color channel image resulting in edge images. The edges are detected horizontally, vertically and both combined.
- Step 3: Gray Level Co-occurrence Matrix (GLCM) is calculated for each edge, holding the features of the edge image.
- Step 4: These features are given to the Support Vector Machine (SVM) to decide whether a forgery is detected or not.

2) *Review on the System Algorithm*: Our algorithm assumes that the images are colored as colors encode relevant information and sensitive to lighting condition at the moment of image acquisition. Therefore, it is expected to have homogeneous color distribution in case of image splicing. Unlike the copy-move forgery detection, we use  $YC_bC_r$  color model instead of gray-scale images.  $Y$  is the component of luminance that contains most of the image content.  $C_b$  and  $C_r$  are the component of chroma blue-difference and red-difference [28].

Our algorithm for image splicing detection works as follows:

*Image Edge detection*: There are multiple edge detector techniques such as Sobel, LoG or Canny. In this paper we adopted similar technique to [27]. We used the edge detection on the equivalent integral image of the input image. We used four edge images which are: vertical, horizontal, diagonal



Figure 8. An example of spliced image and the Diagonal Edge detection of RGB,  $Y$ ,  $C_b$  and  $C_r$  images from top to down and from left to right respectively

and the opposite diagonal which we call the co-diagonal. After obtaining the  $C_r$ , we built Haar-like wavelet filters to find vertical and horizontal edges in the  $C_r$  image. Next, we calculated the integral image, and built a Haar-like wavelet filter, thus, we could construct the vertical and horizontal edges of the image. For the diagonal and the co-diagonal images, we applied the same method, however, a rotated version of the integral image was used instead of the original one.

*Gray Level Co-occurrence Matrix (GLCM)*: After constructing the  $C_r$  edge images, Gray Level Co-occurrence Matrix (GLCM) was applied for texture extraction for each horizontal, vertical, diagonal and co-diagonal edge image. Texture extraction is the equivalent process to the image extraction feature in the copy-move forgery detection. Thus, Texture features are needed to decide the forgery. The Gray Level Co-occurrence Matrix (GLCM) is calculated by creating 8x8 matrix that contains all the features needed for the four edge images. The combination of these matrices generates a feature vector of length 256. This vector will be fed to the classifier for the forgery detection.

3) *SVM Classifier*: Support Vector Machine (SVM) is an efficient and optimal classifier commonly used with machine learning systems, and neural networks [28], [2]. In our system we only have two classes original and fake. So, our model predicts the labels or the classes of our tested features.

4) *Datasets used*: We used CASIA datasets [1] for image splicing, which was divided into two versions; **CASIA I** that consists of 1,737 images (816 authenticated images and 921 spliced images). **CASIA II** consists of 12,625 images (7,492 authenticated images and 5,133 spliced images). We randomly

selected 500 authenticated images and 448 spliced images from both datasets to train and test model. We divided the chosen images into 2 classes; training class (790 images; 417 original images and 373 spliced images), and the testing class (158 images; 83 original images and 75 fake images). An examples for this dataset is shown in Figure 10. Finally, we were limited to colored images as our algorithm works on the  $YC_bC_r$  image components.



Figure 9. An example from Benchmark dataset. The original image is on the left and its fake copy is on the right



Figure 10. An example from CASIA dataset. The original image on the left and its spliced image on the right

#### IV. EXPERIMENTS & RESULTS

In this section, our results are presented for the copy-move and compared with [8], and the same for image splicing compared with [28]

##### A. Copy-Move Results

We examined two different versions of key-points based feature vectors; SIFT and SURF. methods in our system to extract the features from each block to detect identical features and thus, the type of forgery. We compared the SIFT with the SURF to find out which one is better for feature extraction. We ran our algorithm on 3 datasets MICC-F8multi, MICC-F220 [14], and Benchmark datasets [8] as shown in Tables I. From the table it appears that SURF produced more robust results as the number of matched feature points in all test datasets are relatively high when compared to that points matched and were extracted by SIFT.

Table I  
AVERAGE NUMBER OF MATCHING POINTS PER IMAGE

	Datasets		
	Average Matching Points		
	MICC-F8Multi	MICC-F220	Benchmark
SIFT	58	40	1774
SURF	113	120	2023

In Table II the confusion matrix is presented for the Benchmark datasets [8] and MICC datasets [14] with 163 tampered images and 110 original images as shown in Table II.

Table II  
CONFUSION MATRIX FOR COPY-MOVE DATASET

		Predicted	
		Original	Fake
Actual	Original	100	10
	Fake	10	153

Regarding the F-Measurements the achieved True-Positive (TP) rate is 56%, and the False-Negative(FN) rate is 3.8%. The other two metrics: the True-Negative (TN) rate is 36.6% and the False-Positive(FP) rate were 3.6%. The accuracy is 92.67%.

We compared our Benchmark dataset results with [8] as our work on is based on. The results showed that our execution time is less for each mentioned step leading to a decrease in the average execution time for image tampering detection.

According to [8] the average execution time for copy-move detection per image using SIFT is 610.96 seconds, while using SURF is 1052.12 seconds. For our proposed approach, the average execution time using SIFT is 150.8449648 seconds, and for SURF is 89.4841087 seconds. Our Results shows that "Multi-blocking" can enhance the execution time. In addition, it shows that SURF as a feature extractor is more reliable than using SIFT.

##### B. Image Splicing Results

We collected 158 images to test our system, 83 original images, and 75 spliced images. The system converts the input images to  $YC_bC_r$  to detect image splicing. In the following subsection, our results for each image component are presented including the accuracy and performance, beside highlighting the component that gave the best result.

1) *Y Image Component*: We created GLCM on the  $Y$  image component for all images in the dataset. Then we created a training model and added the test feature vectors for all 158 images in the  $Y$  image component. There was 40% fake images detected, which means 30 images out of the 75 fake images were correctly detected. On the other hand, 60% of fake images were falsely detected as original. Also, 80 images of 83 original images were correctly defined as original images. So, the percentage of original images falsely detected as fake images was 3%.

2)  *$C_b$  Image Component*: Again we developed the feature vector for  $C_b$  image component. The results were much better than the  $Y$  image component. The system showed 47% of fake images, which means that 35 images out of 75 spliced images were correctly found. While, the rest of the spliced images 53% were falsely considered as original images which is equivalent to 40 images of 74 spliced images. Regarding the original images 71 images were positively detected from 83 original images. However, there was 14% of original

images falsely detected as fake.

3)  $C_r$  Image Component: Our system gave the best result for  $C_r$  image component in image splicing detection. In Table III we present the confusion matrix of  $C_r$  image component based on 158 images from CASIA dataset [1]

Table III  
CONFUSION MATRIX FOR SPLICING DATASET

		Predicted	
		Original	Fake
Actual	Original	59	24
	Fake	1	74

The results of the  $C_r$  component show that we achieved True-Positive (TP) rate about 99%, and True-Negative (TN) rate greater than 71%. The False-Positive (FP) rate is 29%, and the False-Negative (FN) rate is just 1%. According to [28]  $C_r$  component showed accuracy up to 90.5% which is less than our result by 8.5%.

## V. DISCUSSION

In this section, the results will be discussed and compared to [8] for the copy-move, and [28] for the image splicing. Also, some limitations of the system are discussed.

Our algorithm showed that SURF in extracting features is more reliable than SIFT. According to our results, SURF managed to extract more reasonable matching points from the image blocks, which in return increased the accuracy of detecting the forgery in more than SIFT. Beside, SURF can detect the scaled and rotated forged objects.

In image splicing, we worked with the  $Y$ ,  $C_b$ , and  $C_r$  components individually.  $C_r$  proved its reliability in detecting the splicing higher than  $C_b$  and  $Y$  components.

There are some limitations in our system. First, There were few features extracted in the copy-move algorithm from some of the images in the dataset using our 2 feature extraction methods; SIFT and SURF. One proposed solution can be using another feature extraction as block-based methods such as DCT [24] or DWT [3]. Also, our algorithm depends on dividing the images into blocks in the copy-move detection, however some objects can be divided between multiple blocks which can cause negatively affects the matching point step that compares the features of the blocks to one another.

Concerning the splicing forgery detection, some of edges in integral images were not clear enough to be detected and added to the feature vector of the image. Thus, we propose using combined features instead. Also using a different kernel in the SVM model could be used instead Gaussian or Radial Basis Function (RBF) such as Linear, Polynomial or Sigmoid.

## VI. CONCLUSION

In this work, we presented a general framework for detecting two challenging forgery techniques, the copy-move and splicing. In particular, our system can detect the manipulated regions in the image. Our results show that a key-point based method based on the SURF features, can be more efficient for copy-move forgery detection than SIFT. Its main advantage is the remarkably low computational load, combined with good performance and detection of scaled or rotated objects. We also quantified the performance of splicing forgery detection using SVM model with RBF kernel, which give outstanding results when applied on the  $C_r$  component of the image. We hope our work can serve as an initial building block to improve the security of images on the web. We also believe that our insights would help the forensics professionals with a more concrete decisions.

## REFERENCES

- [1] CASIA V1,II, author=Jing Dong,Wei Wang,Tieniu Tan, howpublished = <https://www.kaggle.com/sophatvathana/casia-dataset>.
- [2] Tim Adams, Jens Dörpinghaus, Marc Jacobs, and Volker Steinhage. Automated lung tumor detection and diagnosis in ct scans using texture feature analysis and svm. In *FedCSIS Communication Papers*, 2018. doi: 10.15439/2018F176.
- [3] Maryam Nabil Al-Berry, Mohammed A.-M. Salem, Hala Mousher Ebeid, Ashraf S Hussein, and Mohammed F Tolba. Fusing directional wavelet local binary pattern and moments for human action recognition. *IET Computer Vision*, 10(2):153–162, 2016.
- [4] Amani A Alahmadi, Muhammad Hussain, Hatim Aboalsamh, Ghulam Muhammad, and George Bebis. Splicing image forgery detection based on dct and local binary pattern. In *2013 IEEE Global Conference on Signal and Information Processing*, pages 253–256. IEEE, 2013. doi: 10.1109/GlobalSIP.2013.6736863.
- [5] Hesham A Alberry, Abdelfatah A Hegazy, and Gouda I Salama. A fast sift based method for copy move forgery detection. *Future Computing and Informatics Journal*, 3(2):159–165, 2018. doi:10.1016/j.fcij.2018.03.001.
- [6] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (surf). *Computer vision and image understanding*, 110(3):346–359, 2008.
- [7] Sevinc Bayram, Husrev Taha Sencar, and Nasir Memon. A survey of copy-move forgery detection techniques. pages 538–542, 2008. doi: 10.1109/ICISC.2017.8068703.
- [8] Vincent Christlein, Christian Riess, Johannes Jordan, Corinna Riess, and Elli Angelopoulou. An evaluation of popular copy-move forgery detection approaches. *IEEE Transactions on information forensics and security*, 7(6):1841–1854, 2012. doi:10.1109/TIFS.2012.2218597.
- [9] A Jessica Fridrich, B David Soukal, and A Jan Lukáš. Detection of copy-move forgery in digital images. In *in Proceedings of Digital Forensic Research Workshop*. Citeseer, 2003. doi:10.1016/j.forsciint.2013.05.027.
- [10] Fahime Hakimi, Mahdi Hariri, and Farhad GharehBaghi. Image splicing forgery detection using local binary pattern and discrete wavelet transform. In *2015 2nd International Conference on Knowledge-Based Engineering and Innovation (KBEI)*, pages 1074–1077. IEEE, 2015. doi:10.1109/KBEI.2015.7436195.
- [11] Tu K Huynh, Khoa V Huynh, Thuong Le-Tien, and Sy C Nguyen. A survey on image forgery detection techniques. In *The 2015 IEEE RIVF International Conference on Computing & Communication Technologies-Research, Innovation, and Vision for Future (RIVF)*, pages 71–76. IEEE, 2015. doi: 10.1109/RIVF.2015.7049877.
- [12] Tu Huynh-Kha, Thuong Le-Tien, Synh Ha-Viet-Uyen, Khoa Huynh-Van, and Marie Luong. A robust algorithm of forgery detection in copy-move and spliced images. *IJACSA International Journal of Advanced Computer Science and Applications*, 7(3), 2016. doi: 10.14569/IJACSA.2016.070301.

- [13] Tu Huynh-Kha, Thuong Le-Tien, Synh Ha-Viet-Uyen, Khoa Huynh-Van, and Marie Luong. A robust algorithm of forgery detection in copy-move and spliced images. *IJACSA International Journal of Advanced Computer Science and Applications*, 7(3), 2016.
- [14] R. Caldelli A. Del Bimbo G. Serra. I. Amerini, L. Ballan. A sift-based forensic method for copy-move attack detection and transformation recovery. pages pp. 1099–1110. *IEEE Transactions on Information Forensics and Security*, vol. 6, issue 3, 2011. doi: 10.1109/TIFS.2011.2129512.
- [15] Pravin Kakar and N Sudha. Exposing postprocessed copy–paste forgeries through transform-invariant features. *IEEE Transactions on Information Forensics and Security*, 7(3):1018–1028, 2012.
- [16] Yongzhen Ke, Fan Qin, Weidong Min, and Guiling Zhang. Exposing image forgery by detecting consistency of shadow. *The Scientific World Journal*, 2014, 2014. doi:10.1155/2014/364501.
- [17] Shinfeng D Lin and Tszan Wu. An integrated technique for splicing and copy-move forgery image detection. In 2011 4th International Congress on Image and Signal Processing, volume 2, pages 1086–1090. IEEE, 2011. doi: 10.1109/CISP.2011.6100366.
- [18] Shinfeng D Lin and Tszan Wu. An integrated technique for splicing and copy-move forgery image detection. In 2011 4th International Congress on Image and Signal Processing, volume 2, pages 1086–1090. IEEE, 2011.
- [19] Tony Lindeberg. Scale invariant feature transform. 2012.
- [20] Cecilia Pasquini, Carlo Brunetta, Andrea F Vinci, Valentina Conotter, and Giulia Boato. Towards the verification of image integrity in online news. pages 1–6, 2015. doi: 10.1109/ICMEW.2015.7169801.
- [21] Christian Riess, Mathias Unberath, Farzad Naderi, Sven Pfaller, Marc Stamminger, and Elli Angelopoulou. Handling multiple materials for exposure of digital forgeries using 2-d lighting environments. *Multimedia Tools and Applications*, 76(4):4747–4764, 2017. doi: 10.1007/s11042-016-3655-0.
- [22] Seung-Jin Ryu, Matthias Kirchner, Min-Jeong Lee, and Heung-Kyu Lee. Rotation invariant localization of duplicated image regions based on zernike moments. *IEEE Transactions on Information Forensics and Security*, 8(8):1355–1370, 2013.
- [23] M.A.-M. Salem. Multi-stage localization given topological map for autonomous robots. In *International Conference on Computer Engineering and Systems, ICCES 2012*, pages 55–60, 2012.
- [24] Mohammed A.-M. Salem, Markus Appel, Frank Winkler, and Beate Meffert. Fpga-based smart camera for 3d wavelet-based image segmentation. In *2008 Second ACM/IEEE International Conference on Distributed Smart Cameras*, pages 1–8. IEEE, 2008.
- [25] BL Shivakumar and S Santhosh Baboo. Detection of region duplication forgery in digital images using surf. *International Journal of Computer Science Issues (IJCSI)*, 8(4):199, 2011.
- [26] Ira Tuba, Eva Tuba, and Marko Beko. Digital image forgery detection based on shadow texture features. In *2016 24th Telecommunications Forum (TELFOR)*, pages 1–4. IEEE, 2016. doi: 10.1109/TELFOR.2016.7818875.
- [27] Paul Viola and Michael J. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001. Volume: 1*, pp.511–518., pages 1257–1260. IEEE, 2009. doi: 10.1109/CVPR.2001.990517.
- [28] Wei Wang, Jing Dong, and Tieniu Tan. Effective image splicing detection based on image chroma. In *2009 16th IEEE International Conference on Image Processing (ICIP)*, pages 1257–1260. IEEE, 2009. doi: 10.1109/ICIP.2009.5413549.

# Weighted Multimodal Biometric Recognition Algorithm Based on Histogram of Contourlet Oriented Gradient Feature Description

Xinman Zhang

School of Electronics and  
Information Engineering  
MOE Key Lab for Intelligent  
Networks and Network Security  
Xi'an Jiaotong University  
No.28 xianning west road  
Xi'an, P.R.China  
zhangxinman@mail.xjtu.edu.cn

Dongxu Cheng

School of Electronics and  
Information Engineering  
MOE Key Lab for Intelligent  
Networks and Network Security  
Xi'an Jiaotong University  
No.28 xianning west road  
Xi'an, P.R.China  
dxcheng@stu.xjtu.edu.cn

Xuebin Xu

Guangdong Xi'an Jiaotong  
University Academy.  
No. 3, Daliangdesheng East Road  
Foshan, China  
ccp9999@126.com.

**Abstract**—Although the unimodal biometric recognition (such as face and palmprint) has higher convenience, its security is also relatively weak. The recognition accuracy is easy affected by many factors such as ambient light and recognition distance etc. To address this issue, we present a weighted multimodal biometric recognition algorithm with face and palmprint based on histogram of contourlet oriented gradient (HCOG) feature description. We employ the nonsubsampling contour transform (NSCT) to decompose the face and palmprint images, and the HOG method is adopted to extract the feature, which is named as HCOG feature. Then the dimension reduction process is applied on the HCOG feature and a novel weight value computation method is proposed to accomplish the multimodal biometric fusion recognition. Extensive experiments illustrate that our proposed weighted fusion recognition can achieve excellent recognition accuracy rates and outmatches the unimodal biometric recognition methods.

## I. INTRODUCTION

WITH the continuous progress and development of the artificial intelligence technology, the conventional recognition technology is unable to meet people's advanced needs. In recent years, biometric identity recognition technology [1] has been widely studied to accomplish the user's identity recognition quickly and accurately.

Since the face biometric recognition technology has higher security and accuracy, it has attracted a large number of scholars to carry out research [2-3]. Many feature extraction methods have been proposed, such as the local binary pattern (LBP) [2], discrete wavelet transform (DWT) [3], nonsubsampling contour transform (NSCT) [4], Histogram of Oriented Gradient (HOG) [5] and Gabor transform [6] et al. These features have also been widely used in palmprint recognition and achieved good results [7-8]. However, when the biometric is affected by the environment factors, the recognition accuracy will decline. To address this issue, some scholars proposed the multimodal biometric recognition theory and carry out extensive research [9-10]. Since NSCT can capture the smoothness and continuities along the contour of the image, it possesses the properties of shift invariance and multi-scale and has been widely studied and employed to describe the features of the face and palmprint [11].

Inspired by this, we propose a novel weighted fusion algorithm based on NSCT and HOG to realize the multimodal biometric recognition with face and palmprint. Firstly, NSCT are utilized to decompose the face and palmprint images. Then, we propose a novel feature extraction method named as histogram of contourlet oriented gradient (HCOG). With full consideration of the intra-class similarity and the inter-class similarity of sample features, a weighted fusion strategy is proposed to realize multimodal biometric fusion recognition.

The main content structure of this paper is organized as follows. Section II discusses the proposed algorithm in detail. In section III, the proposed algorithm is verified by extensive experiments, and the experimental results are analyzed and discussed in depth. Finally, we summarize the conclusion in section IV.

## II. PROPOSED ALGORITHM

### A. Nonsubsampling Contourlet Transform

Contourlet transform [12] employs the Laplace pyramid (LP) transform to decompose the image, then combines the coefficient points with uniform direction into an image contour by using the dimensional filter bank (DFB). However, in the process of image decomposition and reconstruction, contourlet transform needs to adopt the down-sampling and up-sampling operations, which will give rise to frequency aliasing and cause image translation, larger distribution of decomposition coefficient, translation variability and some other problems. To address these issues, Cunha et al. [13] proposed the NSCT algorithm which adopted multi-resolution decomposition by nonsubsampling pyramid (NSP) and nonsubsampling directional filter bank (NSFB) decomposition. Figure 1 illustrates the principle of NSCT [13]. NSCT does not require down-sampling after filtering and up-sampling before band-pass filtering, it only up-sampling on the corresponding filter, and then using low-pass filter and band-pass filter to complete the low-frequency and a series of high-frequency sub-bands filtering.

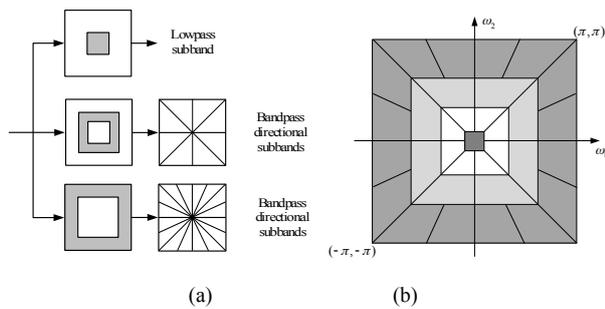


Fig. 1 Nonsampled contourlet transform. (a) NSFB structure that implements the NSCT. (b) Idealized frequency partitioning obtained with the proposed structure.

Since NSCT adopts the nonsampled operation, it has the stability with image shift, that is to say, it has the shift invariance.

### B. Histogram of Oriented Gradient

Histogram of oriented gradient [5] is a classical feature description algorithm, which can describe the edge and shape information of an image well. The specific steps of HOG feature extraction algorithm can be summarized as follows, and figure 2 shows the schematic diagram.

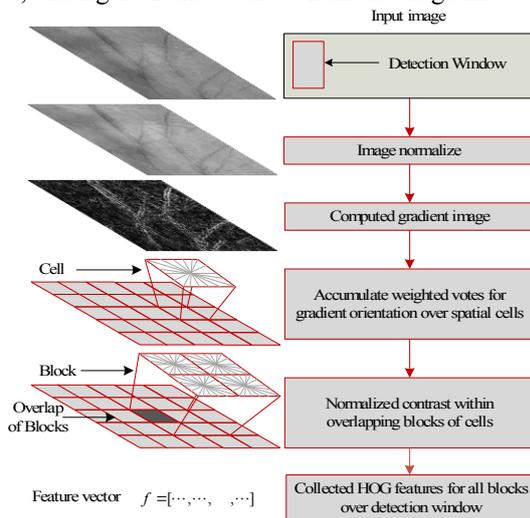


Fig. 2 Schematic diagram of HOG feature extraction.

(1) Normalize the image and calculate the gradient. (2) Divide the image into small cells with the same size and accumulate the weighted votes for gradient orientation with each cell to obtain the histogram of oriented gradient. (3) Group cells into large blocks and use the results to normalize all of the cells in the block. HOG descriptor is obtained by concatenating the normalized histogram vectors of all cells in each block. (4) Tile the detection window with a dense (overlapping is used) grid and cascade the feature descriptor of each block to generate HOG feature vector of the image.

The essence of HOG is the statistical information of image gradient, and the significant position of the gradient is mainly concentrated in the edge of the object. Therefore, it can achieve better recognition effect in the fields of pedestrian detection, and face recognition etc.

### C. Histogram of Contourlet Oriented Gradient (HCOG) Feature Description

Although HOG has achieved stable and good effects in recognition, it lacks multi-scale adaptability due to the restriction of hierarchical rules. To address this issue, Bosch et al. [14] proposed a spatial pyramid representation to encode the object shapes. Inspired by this, we propose a novel HCOG feature descriptor by combining NSCT and HOG to improve the multi-scale performance of the conventional HOG feature.

HCOG feature extraction method is summarized as follows. (1) Implement the multi-scale decomposition of images based on NSCT. Figure 3 illustrates the NSCT decomposition results of a palmprint image, where the decomposition scale is 2, the corresponding bandpass directional subbands numbers are 2 and 4 respectively. (2) Extract HOG feature with the obtained lowpass subband and bandpass directional subbands respectively. (3) Concatenate these HOG descriptors with different subbands and scales to obtain the final HCOG feature vector.

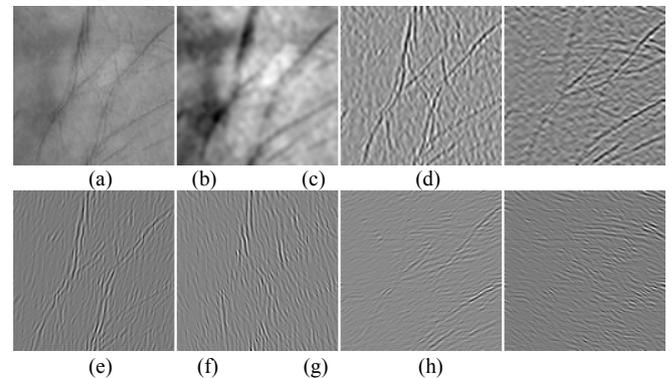


Fig. 3 The 2-level NSCT decomposition demonstration of a palmprint image. (a) Original image. (b) Lowpass subband. (c) and (d) are two different bandpass directional subbands of level 1. (e)-(h) are four different bandpass directional subbands of level 2.

### D. Weighted multimodal biometric recognition algorithm

Although the feature extraction methods of face and palmprint images are same, the feature descriptions of different biometrics are quite different. In order to avoid the mutual interference between different biometrics, we present a novel weighted fusion recognition algorithm with face and palmprint. The specific algorithm steps are as follows.

Step 1. Image preprocessing. Preprocess the training face and palmprint images, including image normalization and ROI extraction etc.

Step 2. NSCT decomposition. NSCT decomposition is performed on the preprocessed face and palmprint images.

Step 3. HCOG feature extraction. Extract HOG feature from subbands obtained by NSCT decomposition with face and palmprint images respectively. Concatenate the HOG descriptors of the NSCT decomposition subbands with each face and palmprint sample to obtain the HCOG feature vector respectively. Then the feature matrix can be obtained, denoted as  $H^f$  and  $H^p$  for face and palmprint respectively.

Step 4. Feature dimension reduction. Employ the linear discriminant analysis (LDA) or principle component analysis (PCA) to reduce the dimensionality of the feature

matrices. Denote the dimension reduced training sample matrices as  $Tr^f$  and  $Tr^p$  for face and palmprint respectively.

Step 5. Weight value calculation. Denote the training matrices as  $Tr^j = [Tr_1^j, Tr_2^j, \dots, Tr_C^j]$ ,  $j = \{f, p\}$ , where  $C$  is the class number, suppose there are  $n$  samples for each class, then the training sample number can be calculated by  $N = nC$ . Denote  $Tr_i^j = [tr_{(n-1)\times i+1}^j, \dots, tr_{n\times i}^j]$ ,  $j = \{f, p\}$  as the training subsets in the  $i$ th class. For each training sample  $x_k^f$  and  $x_k^p$ , calculate the similarity metric between different samples by using  $S_{kl}^j = S(x_k^j, x_l^j)$ ,  $j = \{f, p\}$ , ( $l = 1, \dots, k-1, k+1, \dots, N$ ). Then we calculate the intra-similarity as follows

$$\begin{aligned} \text{intra\_}S^j &= \frac{1}{C} \sum_{i=1}^C \text{intra\_}S_i^j \\ &= \frac{1}{n(n-1)C} \sum_{i=1}^C \sum_{k=(n-1)\times i+1}^{n\times i} \sum_{\substack{l=1 \\ l \neq k}}^{n\times i} S_{kl}^j, j = \{f, p\}. \end{aligned} \quad (1)$$

Calculate the inter-similarity as follows

$$\begin{aligned} \text{inter\_}S^j &= \frac{1}{C} \sum_{i=1}^C \text{inter\_}S_i^j \\ &= \frac{1}{C(C-1)n^2} \sum_{i=1}^C \sum_{k=(n-1)\times i+1}^{n\times i} \sum_{\substack{l=1 \\ l \notin I_i}}^N S_{kl}^j, j = \{f, p\}. \end{aligned} \quad (2)$$

where,  $I_i = \{(n-1)\times i+1, (n-1)\times i+2, \dots, n\times i\}$ ,  $i = 1, \dots, C$ .

Define the discriminant function as follows

$$d^j = \frac{|\text{intra\_}S^j - \text{inter\_}S^j|}{\text{intra\_}S^j + \text{inter\_}S^j}, j = \{f, p\}. \quad (3)$$

Then the weighted value is calculated by

$$\omega^j = \frac{d^j}{d^f + d^p}, j = \{f, p\}. \quad (4)$$

Step 6. Fused recognition. For any given test images of face and palmprint, denoted as  $y = [y^f, y^p]$ , we employ the step 1 to step 4 to obtain the feature vectors, and denote it as  $Tt = [tt^f, tt^p]$ . Calculate the similarity between the test sample and each sample in the training set by  $S_i^j = S(tt^j, tr_i^j)$ ,  $j = \{f, p\}$ , ( $i = 1, 2, \dots, N$ ). Then we use the weighted value calculated by formula (4) to complete the fusion process

$$S_i = \omega^f S_i^f + \omega^p S_i^p, (i = 1, 2, \dots, N). \quad (5)$$

Finally, the nearest neighbor (NN) method is used to implement the recognition task.

### III. EXPERIMENT AND ANALYSIS

In order to verify the effectiveness of the proposed algorithm, we carry out extensive simulation experiments on the multimodal database composed by the face and palmprint.

#### A. Experiment Database

Extended YaleB database contains 38 classes of face images, each class includes 64 different lighting conditions images. After manual cropping, the image size was adjusted to  $192 \times 168$ , and some face images are shown in figure 4.



Fig. 4 Face images of the Extend YaleB database.

Because the red blood cells in the human body has the characteristics of absorbing infrared light in specific bands. This can reduce the influence to the acquired palmprint images caused by palmprint desquamate, injury and sweating etc. We select the palmprint image acquired under near-infrared (NIR) illumination condition to fuse with the face biometric. The biometric research centre of Hong Kong polytechnic university (PolyU) has provided a multispectral palmprint database [1], which mainly includes four different spectral conditions (i.e. blue, red, green and NIR). It contains 500 different classes, and each class contains 12 palmprint images with four different spectral conditions. Figure 5 lists some palmprint images of this database.



Fig. 5 Some NIR palmprint images of the PolyU multispectral palmprint database.

#### B. Experiment Results and Analysis

Here we use the face and NIR palmprint images in Extend YaleB and PolyU database to constructed a multimodal database and conduct the fusion recognition experiment. According to the numbers of sample class and samples contained in each class, we construct a multimodal database with 38 classes and 12 samples for each class. We vary the training sample number from 2 to 7 from each class to demonstrate the experiments, the recognition rates of face and NIR palmprint are tested respectively to compare with the proposed fusion recognition algorithm. In this experiment, we used LDA method to conduct dimension reduction treatment. The specific experimental results are shown in table I.

TABLE I.  
RECOGNITION RATES FOR DIFFERENT BIOMETRICS WITH THE TRAINING SAMPLE VARIES FROM 2 TO 7.

Biometrics	Recognition rates (%)					
	2	3	4	5	6	7
Face	84.74	89.47	93.75	94.74	95.18	97.89
Palm	98.95	98.83	100	100	100	100
Fusion case	99.74	99.71	100	100	100	100

From the table I, it is easy to see that the proposed algorithm outperforms the unimodal biometric recognition cases. Especially, when the number of training samples is small, (such as 2 and 3), our fusion recognition algorithm is 15% and 10.24% higher than the face recognition, and 0.79% and 0.88% higher than NIR palmprint recognition, respectively.

In order to verify the efficiency of our proposed HCOG feature, we implement experiment with different feature extraction methods, such as DWT, LBP, HOG, NSCT and Gabor feature etc. In addition, we use LDA and PCA methods to implement the dimension reduction and compare them with the case by using the original feature. In this experiment, the number of training samples is 3, and the rest are treated as test samples. Specific experimental results are shown in table II.

TABLE II.  
RECOGNITION RATES WITH DIFFERENT FEATURE EXTRACTION METHODS

Feature extraction method	Recognition rates (%)		
	Original feature	LDA	PCA
DWT	97.66	98.54	97.08
LBP	97.95	98.54	97.37
HOG	92.40	91.81	94.74
NSCT	97.08	94.15	96.78
Gabor	98.25	98.54	99.12
Proposed method	99.74	99.74	99.71

From the data in table II, it is easy to see that the feature extraction algorithm proposed by us can effectively improve the recognition accuracy compared with the other feature extraction methods.

In order to verify the efficiency of the proposed algorithm, we draw the cumulative match characteristic (CMC) curve. As illustrated in the figure 6, the CMC curve takes Rank as the abscissa and the cumulative match score as the ordinate, which reflects the identity recognition ability of the biometric recognition system. It can be seen from the figure 6 that our proposed weighted fusion recognition algorithm can always achieve better fusion matching scores, which verifies that our recognition algorithm significantly outperforms the unimodal biometric recognition cases with face or NIR palmprint.

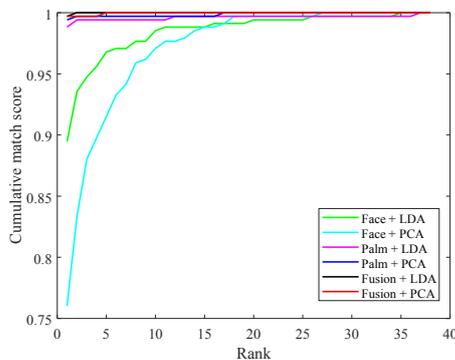


Fig. 6 Performance curves of our proposed weighted algorithms with the unimodal recognition in terms of Cumulative Match Characteristic.

#### IV. CONCLUSION

In this paper, a novel weighted fusion multimodal biometric recognition algorithm with face and palmprint is presented. By integrating the advantages of NSCT and HOG, a

novel HCOG feature descriptor is proposed. At the same time, a fusion weight calculation strategy based on the sample similarity is presented to realize the biometric fusion recognition. Experimental results show that the proposed HCOG feature description and weighted fusion strategy can effectively improve the fusion recognition accuracy.

#### ACKNOWLEDGMENT

This work is supported by National Natural Science Foundation (No. 61673316), Major Science and Technology Project of Guangdong Province (No. 2015B010104002).

#### REFERENCES

- [1] A. K. Jain, A. Ross, and S. Prabhakar. "An Introduction to Biometric Recognition," *IEEE Trans. Circ. Syst. Vid. Tech.*, 2004, vol. 14 pp. 4-20, doi: 10.1109/TCSVT.2003.818349.
- [2] J. Chen, V. Patel, and L. Liu, et al. "Robust Local Features for Remote Face Recognition," *Image and Vision Computing*, 2017, vol. 64, pp. 34-46, doi: 10.1016/j.imavis.2017.05.006.
- [3] A. Ghasemzadeh, H. Demirel. "3D discrete wavelet transform-based feature extraction for hyperspectral face recognition," *IET Biometrics*, 2018, vol. 7, pp. 49-55, doi: 10.1049/iet-bmt.2017.0082.
- [4] X. Xie, J. Lai, and W. Zheng. "Extraction of illumination invariant facial features from a single image using nonsubsampling contourlet transform," *Pattern Recogn.*, 2010, vol. 43, pp. 4177-4189, DOI: 10.1016/j.patcog.2010.06.019.
- [5] N. Dalal, B. Triggs. "Histograms of oriented gradients for human detection," *International Conference on computer vision & Pattern Recognition*, 2005, pp. 886-893, doi: 10.1109/CVPR.2005.177.
- [6] C. Y. Low, A. B. Teoh, and C. J. Ng. "Multi-Fold Gabor, PCA and ICA Filter Convolution Descriptor for Face Recognition," *IEEE Trans. Circ. Syst. Vid. Tech.*, 2019, vol.29, pp.115-128, doi: 10.1109/TCSVT.2017.2761829.
- [7] A. Younesi, M.C. Amirani. "Gabor filter and texture based features for palmprint recognition," *Procedia Comput. Sci.* 2017, vol. 108, pp. 2488-2495, doi: 10.1016/j.procs.2017.05.157.
- [8] S. W. Zhang, H. X. Wang, and W. Z. Huang, et al. "Combining modified LBP and weighted SRC for palmprint recognition," *Signal Image Video Process.* 2018, vol. 12, pp. 1035-1042, doi: 10.1007/s11760-018-1246-4.
- [9] M. Haghghat, M. Abdel-Mottaleb, and W. Alhalabi. "Discriminant Correlation Analysis: Real-Time Feature Level Fusion for Multimodal Biometric Recognition," *IEEE Trans. Inf. Foren. Sec.*, 2016, vol. 11, pp. 1984-1996, doi: 10.1109/TIFS.2016.2569061.
- [10] N. Saini, A. Sinha. "Face and palmprint multimodal biometric systems using Gabor-Wigner transform as feature extraction," *Pattern Anal. Appl.*, 2015, vol. 18, pp. 921-932, doi: 10.1007/s10044-014-0414-6.
- [11] W. F. Li, Y.C. Wang, and Z. Xu Z, et al. "Weighted contourlet binary patterns and image-based fisher linear discriminant for face recognition," *Neurocomputing*, 2017, vol. 267, 436-446, doi: 10.1016/j.neucom.2017.06.045.
- [12] M. N. Do, M. Vetterli. "The Contourlet Transform: An Efficient Directional Multiresolution Image Representation," *IEEE Trans. Image Process.*, 2006, vol. 14, pp.2091-2106, doi: 10.1109/TIP.2005.859376.
- [13] A. L. D. Cunha, J. P. Zhou, and M. N. Do. "The Nonsubsampled Contourlet Transform: Theory, Design, and Applications," *IEEE Trans. Image Process.*, 2006, 15, pp. 3089-3101, doi: 10.1109/TIP.2006.877507.
- [14] A. Bosch, A. Zisserman, and X. Munoz, "Representing shape with a spatial pyramid kernel," *International Conference on Image and Video Retrieval*, 2007, pp. 401-408, doi:10.1145/1282280.1282340.
- [15] D. Zhang, W. K. Kong, and J. You, et al. "Online palmprint identification," *IEEE Trans. Pattern Anal.* 2003, vol. 25, pp. 1041-1050, doi: 10.1109/TPAMI.2003.1227981.

# A GIS Data Realistic Road Generation Approach for Traffic Simulation

Yacine Amara, Abdenour Amamra, Yasmine Daheur, and Lamia Saichi  
Ecole Militaire Polytechnique, Bordj El-Bahri BP 17, Algiers, Algeria  
Email: amara.yacine@gmail.com

**Abstract**—Road networks exist in the form of polylines with attributes within the GIS databases. Such a representation renders the geographic data impracticable for 3D road traffic simulation. In this work, we propose a method to transform raw GIS data into a realistic, operational model for real-time road traffic simulation. For instance, the proposed raw to simulation-ready data transformation is achieved through several curvature estimation, interpolation/approximation, and clustering schemes. The obtained results show the performance of our approach and prove its adequacy to real traffic simulation scenario as can be seen in this video<sup>1</sup>.

**Index Terms**—Road interpolation, Road modeling, Traffic simulation, Vehicle virtual navigation.

## I. INTRODUCTION

In recent years, applications of road traffic simulation have become ubiquitous in everyday life: driving simulation or racing games are increasingly attracting the attention of developers. However, the results obtained are not always consistent with reality. When one wants to reproduce a realistic behavior, the developer must consider the real parameters of the road.

The realization of a real-life road simulation would make it possible to forecast the traffic generated in a given road network. This realism, modeled on a computer, would help managers to detect the problems present in the network, namely congestion, accidents, user stress, an insufficient number of channels, etc. Besides optimizing traffic on the roads without making urban changes, such as location of traffic lights, number, and width of lanes.

The shape of the road is complex, particularly at mountainous segments, severe turns, etc. Therefore, finding the best mathematical function that will fit the shape of these sections while taking into account compressing the number of data to be stored is challenging.

Geometric processing applications rely on the geometric properties of curves such as torsion, curvature, and tangents. In our study, we chose the curvature metric in the road's modeling surface, since the latter embeds the information on the shape of the road. The actual data is extracted from the Geographic Information System (GIS), the latter known for providing the ability to manipulate geographical information laid out on multiple layers. The metric being chosen; our approach is then organized as follows: extraction of the road layer as polylines in the GIS data; curvature estimation at the neighborhood of the points; a grouping of points in clusters according to

the sign of their curvature, and finally the approximation of clusters by continuous mathematical functions.

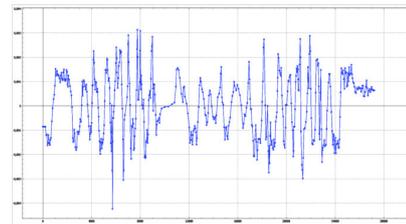


Fig. 1

## II. RELATED WORKS

Most of the work that has been proposed on road reconstruction is based on procedural modeling. Roads are generated by empirical rules often based on observation. We will classify the methods proposed in two categories: the methods taking the input of calculated roads (procedural approaches) and the models taking as input the data resulting from Geographic Information Systems (GIS).

### A. Procedural approaches

These approaches take as input roads derived or calculated by mathematical or optimization methods. Both, Parish et al. [1] and Jing et al. [2] generated models of cities in which road networks played a key role. Their roads were built from the central lines that are generated either by grammar rules named *L-Systems* [1], or model-based methods [2]. However, the defining laws or models to regenerate exactly the structure of urban roads are not available. Hence, their roads often drift from their real respective counterparts. In addition, their methods were not designed for highways and suburban roads.

Some of these limitations have been resolved by Chen et al. [3], where they used a tensor field from which a graph representing the road network was calculated. This can be modified to control locally the generated road profile. Their method permits creating large areas but does not allow the reconstruction of existing networks because of the problem of scale limitation. In addition, only urban roads were considered.

Another procedural model was proposed by Galin et al. [4, 5] based on a platform for generating mapped roads that contain many types of features such as trees, rivers, and lakes. Their roads resulted from a short path algorithm instead of the real data. As a result, we obtain fictitious roads that do not

<sup>1</sup><https://youtu.be/t8eyphcFYHc>

satisfy the constraints of civil engineering. The generated roads are erroneous and do not match the shape of the terrain. This lack of realism is because of the shortcomings of the proposed approaches. The major shortcoming of these works is not to have considered the actual data as input of the problem.

### B. GIS data approaches

To overcome the problems mentioned and to build more realistic roads, Bruneton and Neyret [6] proposed a method of generating large road surfaces from GIS data. They represented the roads with Bezier curves to join the sampled points and then mapped them on the ground. Although this model helps to perceive actual road coordinates, current roads are no longer built from Bezier curves. Again, these models are affected by the lack of realism. The details and constraints related to civil engineering discipline are not taken into account. Road networks must be defined by simple forms such as straight lines, arcs of circles, and clothoids. This drawback had been addressed by the following work:

1) *LSGA algorithm*: This work [7] introduces a new approach to construct smoothed curve pieces representing realistic roads. Given a GIS database of road networks, where the sampled points are organized as 3D polylines, this method creates horizontal and verticals curves, then it combines them to generate the roads. The major contribution of this work is a tree traversal algorithm that extends the sequences of the best fit primitives and a fusion process of these primitives. The latter must respect a certain grammar according to an automaton. This approach offers more realistic results than those that preceded it, and the errors are proportional to the noisiness of input data.

2) *Construction by clothoids*: In [6], the algorithm for adjusting a sequence of  $G_2$  polylines into clothoid segments takes place in two steps: first a piecewise linear approximation is applied, then a sequence of rigid 2D transformations is applied in order to align the in one consistent result. Although this method respects civil engineering constraints and models the transition between a circular arc and a straight line with a clothoid, its first pass through the linear segments loses the precision when estimating the radius of curvature.

3) *Automatic generation of 3D roads*: This method, presented in [8], is based on a set of civil engineering rules. It proposes a new approach for the automatic 3D generation of high-fidelity roads. It transforms GIS data that only contains 2D information from the central axis of the road into a 3D model of the road network. In the proposed approach, basic road elements such as road segments, road intersection are generated automatically to form sophisticated road networks. But in the modeling of the axe of the road, the segments were connected by Hermite curves, which satisfy only  $G_1$  continuity, hence the civil engineering constraints were not respected.

The proposed model must be realistic, that is to say, that it must meet the constraints of civil engineering and vehicle dynamics, for this reason, our study was based on an essential criterion which is *curvature*. Curvature, the inverse of the

radius of the circle tangent to the curve, is defined as the norm of the acceleration vector of a body traveling the curve at unit speed. It is the second derivative with respect to the curvilinear abscissa of the body position.

## III. CURVATURE ESTIMATION

### A. Definition

A parametric curve is a function  $r : I \subset \mathbb{R} \rightarrow \mathbb{R}^n$ , when  $n = 2$  it is called a *plane curve*. The curvilinear abscissa  $s$  from a point  $r(t_0)$ ,  $t_0 \in I$ , at a given point  $r(t_1)$ ,  $t_1 \in I$ , is defined by:

$$s(t_1) = \int_{t_0}^{t_1} \|r'(u)\| du \quad (1)$$

The vector  $T(s) = r'(s)$  is called the tangent vector. The normal vector  $N(s)$  is obtained by a rotation of  $90^\circ$  anticlockwise. The vectors  $T'(s)$  and  $N(s)$  are collinear. That is, there is a function  $k(s)$  such that:

$$T'(s) = k(s) \times N(s) \quad (2)$$

called the *curvature* of the curve at the point  $r(s)$ . The curvature also corresponds to the variation of the direction of the tangent vector respectively to the curvilinear abscissa:  $k(s) = \theta'(s)$ , such that :

$$\theta'(s) = \angle(\overrightarrow{T(s)}, \overrightarrow{(1, 0)}) \quad (3)$$

Since our initial data is in discrete form, we performed a local estimation using a sliding window. The latter is centered around a point allowing to approximate all its neighbors within the window by a second order polynomial. The interest of such an operation is to deduce the first and the second derivatives, to then calculate the curvature. Several approximation methods called *implicit parabola fitting* proposed in [9], which have a good performance and a fair simplicity of implementation.

### B. Second-order curve approximation

In the implicit parabola fitting method, proposed in [9], the curve is described by a function such that:  $y = f(x)$  or  $x = f(y)$ . The variation of  $x$  and  $y$  inside the window determines the parameterization to adopt, in fact if the variation of  $x$  is greater than that of  $y$ , the algorithm will select the case  $y = f(x)$  and vice versa. In order to simplify the notation, we will consider that  $p_0 = (0, 0)$ . The goal is to find  $f'_0$  and  $f''_0$ , which minimize:

$$E_x(f'_0, f''_0) = \sum_{i=-q}^q (y_i - f'_0 x_i - \frac{1}{2} f''_0 x_i^2)^2 \quad (4)$$

The solution of this problem of least squares gives:

$$f'_0 = \frac{cg - bh}{ac - b^2} \quad f''_0 = \frac{ah - bg}{ac - b^2} \quad (5)$$

such that :

$$\begin{aligned} a &= \sum_{i=-q}^q x_i^2 & g &= \sum_{i=-q}^q x_i y_i & b &= \frac{1}{2} \sum_{i=-q}^q x_i^3, \\ h &= \frac{1}{2} \sum_{i=-q}^q x_i^2 y_i & c &= \frac{1}{4} \sum_{i=-q}^q x_i^4 \end{aligned} \quad (6)$$

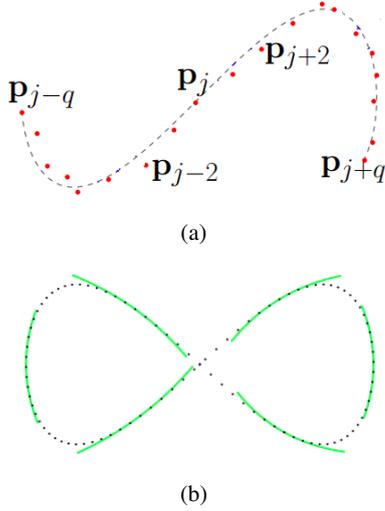


Fig. 2: (a) Points window, (b) Approximation by least squares

### C. Curvature computation

1) *Model and notation:* Consider a set of points  $p_i$  of a flat smooth curve  $r$ , in this study the curve is parameterized by the arc length, the estimate of the curvature for a plane curve requires an approximation of the first and second derivatives of  $r(s)$ . Let a point  $p_0$  be chosen. The derivation of  $r$  in  $p_0$  will be estimated from a window of size  $2q + 1$  around  $p_0 : p_{-q}, p_{-q+1}, \dots, p_q$  (see Fig 2a). We set  $p_0 = r(0)$  as the origin, the approximation of degree two can be written in the form:

$$r(s) = r'(0)s + \frac{1}{2} r''(0)s^2 \quad (7)$$

2) *The least squares approach:* The estimate of  $r'(0)$ ,  $r''(0)$  is obtained by the least squares approach (Fig 2b). The weighting  $w_i$  of the point  $p_i$  must be positive, relatively important for small values of  $|s_i|$  and relatively small for large values of  $|s_i|$ .

The arc length  $s_i$  can be estimated as follows:  $\Delta l_k$  is the arc length of the vector  $p_k p_{k+1}$ , where  $k$  varies from  $-q$  to  $q - 1$ . The arc length between  $p_0$  and  $p_i$  can be approximated by:

$$\begin{cases} l_i = \sum_{k=0}^{i-1} \Delta l_k, & i > 0 \\ l_i = -\sum_{k=0}^{-i-1} \Delta l_k, & i < 0 \end{cases} \quad (8)$$

### D. Curvature computation for a plane curve

For the case of plane curves, the idea of the Independent coordinates method [9] is to construct a parametric curve  $(x(s), y(s))$  that approaches the curve locally, by quadratic functions as a function of arc length.

$$\begin{cases} \hat{x}(s) = x_0 + x'_0 s + \frac{1}{2} x''_0 s^2 \\ \hat{y}(s) = y_0 + y'_0 s + \frac{1}{2} y''_0 s^2 \end{cases} \quad (9)$$

The derivatives  $x'_0$  and  $x''_0$  are estimated by minimizing:

$$E_x(x'_0, x''_0) = \sum_{i=-q}^q w_i (x_i - x'_0 l_i - \frac{1}{2} x''_0 l_i^2)^2 \quad (10)$$

The minimization of this equation can be written in the following matrix form

$$\begin{bmatrix} a_1 & a_2 \\ a_2 & a_3 \end{bmatrix} \begin{bmatrix} x'_0 \\ x''_0 \end{bmatrix} = \begin{bmatrix} b_{x,1} \\ b_{x,1} \end{bmatrix} \quad (11)$$

such that :

$$\begin{aligned} a_1 &= \sum_{i=-q}^q w_i l_i^2 & a_2 &= \frac{1}{2} \sum_{i=-q}^q w_i l_i^3 & a_3 &= \frac{1}{4} \sum_{i=-q}^q w_i l_i^4 \\ b_{x,1} &= \sum_{i=-q}^q x_i w_i l_i & b_{x,2} &= \frac{1}{2} \sum_{i=-q}^q w_i l_i^2 x_i \end{aligned} \quad (12)$$

---

#### Algorithm 1: Weighted least square variables setting

---

$I[] = a_1 = a_2 = a_3 = b_{x,1} = b_{x,2} = b_{y,1} = b_{y,2} = 0$

**for**  $i = -q; i \leq q; q++$  **do**

$I[i] \leftarrow I[i - 1] + \|p_i p_{i-1}\|$

**end**

$m = I[0]$

**for**  $i = -q; i \leq q; q++$  **do**

$I[i] \leftarrow I[i] - m$

$w \leftarrow \text{weight}(I[i])^2$

$a_1 \leftarrow a_1 + w(I[i])^2$

$a_2 \leftarrow a_2 + \frac{w}{2}(I[i])^3$

$a_3 \leftarrow a_3 + \frac{w}{4}(I[i])^4$

$b_{x,1} \leftarrow b_{x,1} + w(I[i])(x_i - x_0)$

$b_{y,1} \leftarrow b_{y,1} + w(I[i])(y_i - y_0)$

$b_{x,2} \leftarrow b_{x,1} + \frac{w}{2}(I[i])^2(x_i - x_0)$

$b_{y,2} \leftarrow b_{y,1} + \frac{w}{2}(I[i])^2(y_i - y_0)$

**end**

$d \leftarrow a_1 a_3 - a_2^2$

---

The same procedure is applied for the calculation of  $y'_0$  and  $y''_0$ . The tangent  $T$  is obtained by the normalization of the vector  $r'_0 = (x'_0, y'_0)$ , while the normal vector is obtained by a rotation of  $90^\circ$  of  $T$ .

## IV. CUTTING ACCORDING TO CURVATURE

This step will aim to cut the road into a set of primitives consisting of straight lines, left/right turns based on the value of curvature estimated at each point as explained in the previous step.

**Algorithm 2:** Coefficient computation

$$\begin{aligned}
x'_0 &\leftarrow (a_3b_{x,1} - a_2b_{x,2})/d \\
y'_0 &\leftarrow (a_3b_{y,1} - a_2b_{y,2})/d \\
x''_0 &\leftarrow (a_1b_{x,2} - a_2b_{x,1})/d \\
y''_0 &\leftarrow (a_1b_{y,2} - a_2b_{y,1})/d \\
\kappa &\leftarrow (x'_0y''_0 - y'_0x''_0) / \|(x'_0, y'_0)\|^3 \\
T &\leftarrow (x'_0, y'_0) / \|(x'_0, y'_0)\| \\
N &\leftarrow \text{sign}(\kappa)(-T_y, T_x)
\end{aligned}$$

For this, we will execute the following processes in the order indicated. We first start with a classification in two primitives only, that is to say left/right turn, we continue by extracting the sections which are straight lines independently of the primitives obtained previously and we thus finish by eliminating the isolated points, which are of the right type because a primitive of the right type must have a minimum number of two points. In what follows we will detail the procedure for filtering the curvature values according to the order of execution of the steps.

*A. Preprocessing*

In order to facilitate the different treatments, an initial marking of all the points of the curve has been carried out. For this, we assigned to each point whose curvature is positive the number +1, and -1 for those whose value of the curvature is negative, and we stored the result in a table named "ids" which will be subsequently updated by the results of the different processing.

*B. Left and right turns detection*

Our approach was based on a local estimate of the curvature sign near the point considered. A sliding window of size  $w$  has been used to estimate the global sign of the values inside the window by summing the values previously assigned, so the classification of the types of sections will be carried out as following :

- The sum is positive, so it is a right turn and the point is marked +1.
- The sum is negative, so it is a left turn and the point is marked -1.
- The zero sum is not a possible case because the size of the window around the point is odd.

This step is very important because it allows to assign a point to a given turn even if the sign of its curvature does not correspond, because it is the trend of the whole neighborhood is taken into consideration.

*C. Straight line detection*

Straight line detection is based on the value of the estimated curvature. Indeed, we know that a line is characterized by a null curvature. Since the values of the estimated curvature are not all accurate and contain noise, a threshold  $\delta$  has been set. If the curvature norm is below this threshold, the point will be assigned to a straight line primitive. Points satisfying the relationship  $|K(s)| < \delta$  will be assigned the number 0.

*D. Elimination of isolated right point*

Knowing that a marked point of a straight line can not be isolated, the solution is to go through all the points of the line and to modify those which are of marking different from their neighbors according to the signs of the curvature at this point, i.e. +1 marking if the curvature is positive, -1 otherwise.

**Algorithm 3:** Primitive assignment

```

while  $i < \text{ids.size}()$  do
  gpts gpt
  gpt.type  $\leftarrow \text{ids}[i]$ 
  gpt.id  $\leftarrow i$ 
  gpt.nbr  $\leftarrow 0$ 
  for  $i = 0; i < \text{ids.size}(); i++$  do
    if  $\text{gpt.type} \neq \text{ids}[i]$  then
      Break
    end
    gpt.nbr++
  end
end

```

## V. TURNS APPROXIMATION WITH SECOND ORDER POLYNOMIALS

*A. Preprocessing*

In order to interpolate curves with polynomials, a data structure has been created to facilitate processing on the one hand and to save the results obtained on the other hand. This data structure ("gpts") consists of an integer field named "type", which will contain the type of the primitive according to the marking carried out in the previous step. An integer field named "id", which will contain the identifier of the first point of the primitive. An integer field named "nbr", which will contain the number of points belonging to this primitive as well as two other fields "paramX" and "paramY", which will be used to store the coefficients of polynomials associated with turns. The structure being created, the fields were subsequently assigned according to algorithm 3.

For each left-handed or right-turn type primitive, we will proceed to the parameterization according to the length of the arc, so we will have two polynomials according to  $x$  and  $y$  such that:

$$\begin{cases} x(s) = a_0 + a_1s + a_2s^2 + a_3s^3 \\ y(s) = b_0 + b_1s + b_2s^2 + b_3s^3 \end{cases} \quad (13)$$

The polynomials in question will be of degree 3, this choice can be justified by the absence of inflection points since the primitives have been classified according to their (there is not a transition in the same road segment). They are convex or concave curves hence the choice of the polynomial regression that is a statistical analysis that describes the variation of an independent random variable, called here  $x$  or  $y$ , according to an independent random variable, called here  $s$ , being the length of the arc. We seek, by regression, to bind the variables by

a polynomial of degree 3. The calculation of the coefficients therefore amounts to solving a system of equations that can be expressed in the following matrix form:

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 1 & s_1^1 & s_1^2 \cdots & s_1^m \\ 1 & s_2^1 & s_2^2 \cdots & s_2^m \\ \vdots & \vdots & \ddots & \vdots \\ 1 & s_n^1 & s_n^2 \cdots & s_n^m \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{pmatrix} \quad (14)$$

The number  $m = 3$  corresponds, in our case, to the cubic polynomial regression. We will treat the case of the polynomial  $x(s)$  generated and it is the same for  $y(s)$ . The problem is therefore to find the vector  $\vec{a}$  in the equation:

$$\vec{x} = S\vec{a} \quad (15)$$

The solution is given by:

$$\vec{a} = (S^T S)^{-1} S^T \vec{x} \quad (16)$$

### B. Error criterion

Given the needs and requirements of the realism imposed on the generated road, we set ourselves the objective of approximating the GIS input data with lines and polynomials of third degrees such that the maximum deviation between the two curves does not exceed 2.0m, in this part we discuss the calculation of error as well as treatments to be undertaken in the case of exceeding this limit.

1) *Error computation*: The error  $d$  is estimated as the distance between the approximate point  $p = (x, y)$  and the entry point  $p = (x, y)$ , it is deduced according to the following formula:

$$d = \sqrt{(\hat{x}(s) - x(s))^2 + (\hat{y}(s) - y(s))^2} \quad (17)$$

The maximum value of this error must not exceed  $\delta$  for each of the points of the different right or polynomial primitives.

$$d_{max} < \delta \quad (18)$$

2) *Error processing*: Primitives whose error exceeds the required threshold must be modified, for this our approach is to generate from the initial primitive two primitives such that the number of points constituting each primitive is equal to half the number of points of the primitive. For a given primitive  $C_m$  one generates two primitives  $C_{m/2}$  and  $C'_{m/2}$ , and the same treatment will be executed on the two new primitives in a recursive way, this will assure us:

- Compliance with the error limit required for each final primitive obtained,
- The generation of a minimal number of primitives,

## VI. RESULTS

The results obtained when applying our method are summarized in Figure (3). The figures show the stages of our approach. Figure (3a) shows a sample of a road polyline. In Figure (3b), the curvatures were estimated by the least squares method. Note that this representation is more exploitable than that of Figure (1), and that there is less disturbance in the values of the curvatures due to windowing.

### A. Curvature estimation by the implicit parabola fitting method

In order to overcome the problems encountered when calculating the curvature by cubic splines, we opted for an estimation of the latter by the least squares. Compared with spline, the quality of curvature values improved significantly. Indeed, most points follow the trend of the turn where they belong as can be seen in Figure (3b). Nevertheless, in the case of more complex turns, outliers appear. Hence, we have proposed to neglect them, and to mark a window by a single sign of curvature relative to reality.

### B. Clustering

Figure (3c) shows that the points of the input road were grouped into three basic categories: left turn, right turn, and straight line. We note that the results are more refined, because this step corrects any possible residual error of the previous section. Moreover, it allows noise reduction, by defining the straight line segments from a given threshold, since the SIG data being noisy, we cannot obtain zero curvature values.

This step allowed us to have the same signs of curvature for a given type of cluster. Nevertheless, at this stage the clusters are not connected to each other, a major problem on which the approximation will be based in the following phases.

To overcome the problem of connection, we fix first the ends of the primitives, then we approximate the calculated model of the initial points. In this perspective, the Bezier curves with a least squares approximation prove to be an adequate choice. Indeed, the first and last control point are superimposed on their correspondents in the initial data, then the other points are calculated by minimizing the differences between the model and the initial data. The main disadvantage of this method is that it guaranties only  $C_0$  continuity between two clusters of successive points, which does not satisfy civil engineering constraints. Moreover, in some complex turns, it remains difficult to follow the shape of the cluster by a third degree polynomial.

Figure (3g, 3h) represents the constructed road. The initial data points circled and clustered. Note that the resulting model does not deviate from the input road polyline. The curves in magentas represent the edges of the road, the input data (polylines) are in red, and the model (at the central axis) is in blue. Figure (3h) illustrates the mapping of the road to the geo-referenced satellite image corresponding to the road section. As can be seen, the model provides a smooth representation of the road surface. We see that complex shapes such as turns are well approximated. In addition, our model meets the  $C_2$  continuity (imposed by road civil engineering), along the road without oscillations or other erratic behaviors that would compromise the visual comfort during the simulation.

## VII. CONCLUSION AND FUTURE WORKS

The objective of this work was the realistic modeling of the road surface by taking into account a number of civil engineering and vehicle dynamics constraints. We chose curvature

as a parameter describing the shape of the road in order to approach the profile of the latter as closely as possible.

We proceed through the estimation of curvature by a windowing approach because of the discrete nature of GIS data. In order to approach the reality, we grouped the points into clusters according to their neighborhood's curvature to obtain basic forms of the road namely left/right turns and straight lines.

This being done, the next step was to find the mathematical functions as well as the appropriate conditions and constraints to approach the initial data with functions that respect  $C_2$  continuity conditions.

The results obtained are satisfactory insofar as our road reconstruction approach takes into account the real constraints, moreover the model obtained is of  $C_2$  continuity, which reflects the smoothness of the position, speed and the acceleration of the simulated vehicle.

As perspective, we plan to improve and refine this work by considering road intersections, since the latter are frequently encountered in real-life situation.

#### REFERENCES

- [1] Yoav IH Parish and Pascal Müller. "Procedural modeling of cities". In: *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. ACM. 2001, pp. 301–308.
- [2] Jing Sun et al. "Template-based generation of road networks for virtual city modeling". In: *Proceedings of the ACM symposium on Virtual reality software and technology*. ACM. 2002, pp. 33–40.
- [3] Guoning Chen et al. "Interactive procedural street modeling". In: *ACM transactions on graphics (TOG)*. Vol. 27. 3. ACM. 2008, p. 103.
- [4] Eric Galin et al. "Authoring hierarchical road networks". In: *Computer Graphics Forum*. Vol. 30. 7. Wiley Online Library. 2011, pp. 2021–2030.
- [5] Eric Galin et al. "Procedural generation of roads". In: *Computer Graphics Forum*. Vol. 29. 2. Wiley Online Library. 2010, pp. 429–438.
- [6] Eric Bruneton and Fabrice Neyret. "Real-time rendering and editing of vector-based terrains". In: *Computer Graphics Forum*. Vol. 27. 2. Wiley Online Library. 2008, pp. 311–320.
- [7] Hoang Ha Nguyen, Brett Desbenoit, and Marc Daniel. "Realistic road path reconstruction from GIS data". In: *Computer Graphics Forum*. Vol. 33. 7. Wiley Online Library. 2014, pp. 259–268.
- [8] Jie Wang, Gary Lawson, and Yuzhong Shen. "Automatic high-fidelity 3D road network modeling based on 2D GIS data". In: *Advances in Engineering Software* 76 (2014), pp. 86–98. DOI: 10.1016/j.advengsoft.2014.06.005. URL: <https://doi.org/10.1016%2Fj.advengsoft.2014.06.005>.
- [9] Thomas Lewiner et al. "Curvature and torsion estimators based on parametric curve fitting". In: *Computers & Graphics* 29.5 (2005), pp. 641–655. DOI: 10.1016/j.cag.2005.08.004. URL: <https://doi.org/10.1016%2Fj.cag.2005.08.004>.

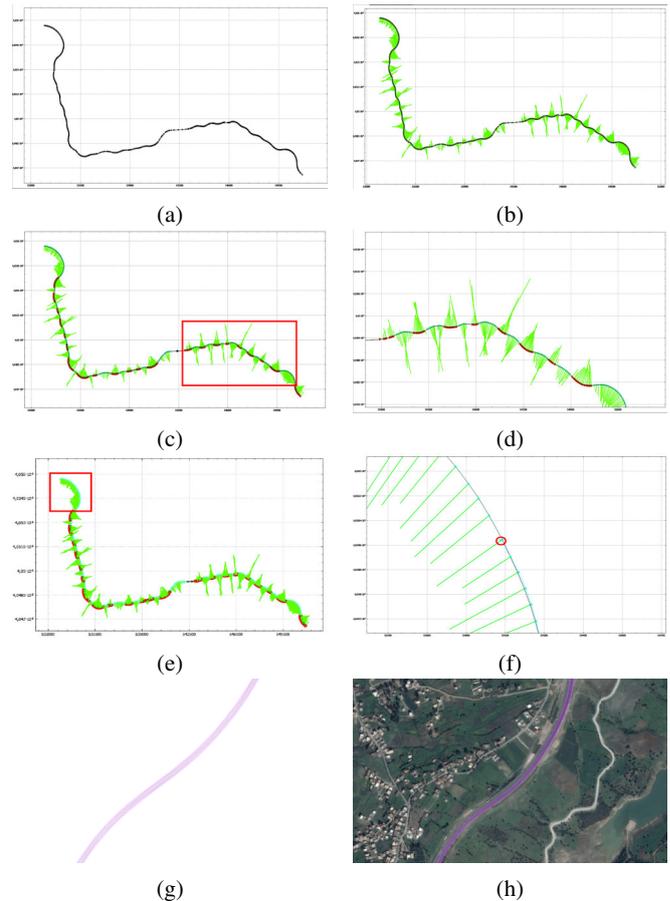


Fig. 3: Processing steps of the proposed algorithms. (a) Raw road in polyline form, (b) Curvature obtained by least squares according to the normal, (c) Cutting of the road (left, right turns and straight lines) according to curvature values, (d) Turns detected by the cutting algorithm, (e) Polynomial fitting of the detected turns, (f) 3rd degree polynomials approximation of a given turn, (g) Result: generated road surface, (h) Mapping of the result on a real road

# Crime Scene Reconstruction with RGB-D Sensors

Abdenour Amamra, Yacine Amara, Khalid Boumaza, and Aissa Benayad

Ecole Militaire Polytechnique, Bordj El-Bahri BP 17, Algiers, Algeria

Email: amamra.abdenour@gmail.com

**Abstract**—Photographic surveying, a fundamental procedure in crime investigation, is typically performed using 2D cameras. Although useful, such cameras remain limited due to the lack of depth information. In this work, we propose a 3D reconstruction solution that leverages the advantages of cheap RGB-D sensors to create a 3D model of the crime scene and to provide the investigator with an interactive crime scenario simulation environment. A structure from motion approach is proposed in order to align the captured point clouds on each other using 3D key points. An iterative refinement and a global optimization algorithm are later adapted for the optimization of the registered 3D model, which is then triangulated before the underlying surface is reconstructed. The resulting model is used for interactive crime investigation and object dynamics simulation. The obtained results show the effectiveness of our solution with a visually appealing rendering, an accurate simulation and a quantitative error of less than 18cm for the  $4m \times 4m$  indoor scene. An accompanying video is provided in order to illustrate the processing pipeline<sup>1</sup>.

**Index Terms**—Crime scene modeling, 3D registration, RGB-D sensors, Forensic computing, Interactive investigation.

## I. INTRODUCTION

In the world we live today, we hear frequently about crimes and the ways they are resolved. Freezing and preserving the crime scene is certainly one of the most important and delicate initial steps that the police perform upon the arrival to a crime scene. The idea is to conserve the spatial configuration of the objects surrounding the place where they think the crime took place for the purpose of avoiding any contamination.

In most current cases, the investigation is carried out in the simplest manner, that is, with paper, pencil and measurement ribbon, to sketch an illustration of the crime scene showing the position of the victim, as well as the other tools that could have served the criminal. For instance, a trivial improvement of the simple flat drawings can be a 3D representation of the scene allowing a better understanding of the series of events that conducted to the crime. For instance, several computer tools have been used by experts for the creation, the visualization, and sharing of electronic crime data among investigators. A first 3D reconstruction method based on this software is the design of a 3D model of the scene. Then taking pictures and pasting them on the 3D mesh. This method is simple and does not require much experience, however, it fails to preserve the 3D shape and dimensions of objects. Another method of 3D scene reconstruction is a faithful recreation of the scene. The latter requires the mastery of sophisticated 3D modeling tools and the measurements of all the important scene objects. Its advantage is that the resulting model is very close to reality.

<sup>1</sup><https://youtu.be/IYnJSNV7QkI>

Unlike the methods mentioned above, our approach is based on the alignment of several 3D images of the same scene, with different points of view. This alignment is looking for spatial transformations that merge the views into a single globally consistent model. With such an approach, the reconstruction of scenes is not left to the skills and performance of a human actor; rather, it relies on real environmental measurements while taking into account the dynamics of objects. Interestingly, with a coherent 3D representation and a physics engine, physical laws can be simulated (e.g. simulation of a bullet's trajectory). Otherwise, it would be possible to test crime hypothesis without the need for deployment on the physical scene. For this purpose, the Kinect could serve as an affordable 3D scanner that has interesting performance in 3D shape capturing. Indeed, RGB-D cameras capture the color and the geometry of the scene and deliver colored 3D point clouds, but their data still needs a chain of preprocessing and 3D registration as well as surface inference in order to become a useful representation. In our case, these sensors were placed at the center of the crime scene and the forensic police will take care of swiveling it in order to scan the objects of interest.

In the remainder of the paper, we first discuss the literature of crime scene reconstruction and modeling in Section II. Then, we present our approach to 3D reconstruction and interaction in Section III. We validate our findings with several quantitative and qualitative assessments in Section IV. And we conclude with a summary of what we aimed for and what we really achieved, as well as some future perspectives that can enhance and extend the present work in Section V.

## II. RELATED WORKS

The attempt to reconstruct crime scenes with 3D reconstruction means finds its origins in [1]; where the authors used a mobile camera in a Structure-from-Motion (SFM) pipeline. Afterward, the authors in [2] presented a general comparison of 3D imaging sensors for criminal investigation. They took into account most of the 3D techniques available at that date. It should be noted, however, that Time of Flight (TOF) technology, the working principle of the Kinect v2, was not considered since the technology has not been widespread until the last decade. The authors in [3] used an alternative dense reconstruction approach directly on video sequences.

Initially used to solve the problem of ego-motion, Simultaneous Localization and Mapping (SLAM) techniques have been a hot topic in robotics for several years. Crime scene reconstruction works had followed through the light of SLAM technology; where the authors of [4] investigated the

utilization of a stereo camera rig in a SLAM based crime scene reconstruction pipeline. Nevertheless, it is worth noting that stereo mapping is a passive technique that reconstructs a cloud of 3D points by matching the corresponding key points found in both RGB images. Since it relies only on the feature points of the environment, RGB imaging techniques are subject to singularities when the images lack distinctive patterns (e.g. uniformly colored surfaces, which potentially characterize indoor scenes; thus, rendering stereo pairing difficult and resulting in a rough and distorted 3D content).

Another approach is to design a 3D model of the scene using 3D modeling software such as 3ds Max [5]. More recent approaches use laser scanners [6] to reconstruct the crime scene at a very high accuracy. An example of the laser scanners that were used for this purpose is FARO S-350 [7]. The latter is dedicated to fast and accurate 3D indoor and outdoor environments. Nonetheless, its high cost and lengthy reconstruction time are two famous shortcomings of such a technology.

In the light of the literature, and in the purpose of building upon the previous contributions, we investigate the utilization of RGB-D data, in an SFM pipeline with a loop closure mechanism and a 3D triangulation and surface reconstruction algorithm for the development of an interactive crime investigation solution.

### III. 3D CRIME SCENE RECONSTRUCTION

In what follows, we present our solution to 3D crime scene reconstruction and interactive investigation.

#### A. Solution Overview

Figure 1a illustrates the process followed to complete our work. As illustrates the diagram, our solution is essentially divided into three steps: a *data acquisition*, where the point clouds are delivered by the Kinect v2 RGB-D sensor. The second step is the *prepossessing* of the point clouds delivered by the sensor. The latter includes: down-sampling, filtering, and key points extraction and description. The last step, which is the most important, is the *3D alignment and reconstruction* that results in a coherent holistic 3D model for investigation purposes.

#### B. Data Acquisition

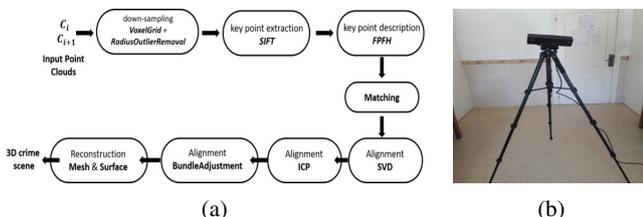


Fig. 1: Crime scene reconstruction setup. (a) Reconstruction pipeline, (b) RGB-D data acquisition setup

This phase involves collecting data from the scene with the Kinect, with the consideration that two successive images

overlap each other. To this end, and as the Kinect sees only the scene within its viewport, we place it at the center of the scene (see Figure 1b) and we have it rotated at small regular angles. We can, therefore, acquire images at multiple views during a complete turn for later reconstruction.

Despite the use of the rotating acquisition procedure, parts of the scene remain occluded in the captured views. In fact, these regions are not visible to the camera as shows Figure 2b. Such a phenomenon can be overcome by adapting the scans to cover the holes. At this level, we focus only on the acquisition of visible parts, because the missing parts can be treated separately after the reconstruction finishes.

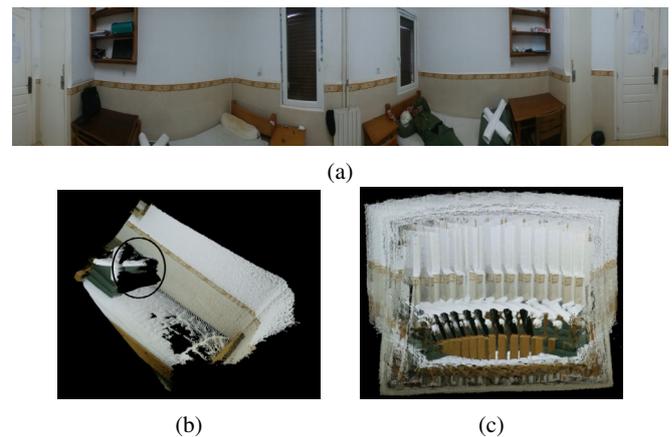


Fig. 2: RGB-D images captured from the experimental  $4m \times 4m$  crime scene. (a) Panoramic view of the crime scene, (b) Point cloud, with hidden regions, obtained from the captured view, (c) Projection of all the views in the same reference frame.

#### C. Filtering and down-sampling

In order to obtain a decent 3D model of the scene, and in the sake of faster 3D mesh creation, it is important to down-sample the raw point cloud. The down-sampling technique that we adapted is based on the Voxelgrid method [8]. After a consistent sampling, we obtain a less dense point cloud that preserves the underlying geometry without redundancy. The second preprocessing to be carried out is noise and outlier points elimination. For this purpose, we adapted the Radius Remove Outlier filter [9]. The latter smoothes the 3D points and removes isolated points.

#### D. Alignment and 3D reconstruction

This section is essential to our contribution, the objective here is the alignment of the different views in order to obtain a coherent 3D scene. Our approach to point cloud alignment is cumulative pairwise. In other words, we align a pair of point sets at a time, then we carry on with the following pair, and so on; until the last view (typically overlapping the first one, as we rotated the camera in order to scan the whole scene from the center).

Technically the alignment of the point sets results from the computation of a  $4 \times 4$  transformation matrix  $T$ . The latter embeds a  $3 \times 3$  rotation matrix and a  $3 \times 1$  translation vector. We add to this matrix the so-called homogeneous coordinates for the simplicity of computation to obtain the  $4 \times 4$  transformation matrix.

$$T = \begin{bmatrix} r_{00} & r_{01} & r_{02} & t_0 \\ r_{10} & r_{11} & r_{12} & t_1 \\ r_{20} & r_{21} & r_{22} & t_2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

For example, in order to align three point clouds  $A_1, A_2, A_3$  together, one first aligns the second cloud  $A_2$  with the first  $A_1$ , where the result after minimization is the transformation  $T_{1,2}$ . We obtain a global intermediate result  $G_{1,2} = A_1 \cup T_{1,2} * A_2$ . Then, we align the third cloud  $A_3$  with the already obtained global result  $G_{1,2}$  to obtain the final result  $G_{1,2,3} = G_{1,2} \cup T_{1,2,3} * A_3$ .

In order to perform point cloud alignment, we propose the following scheme:

#### 1) Initial alignment

This phase aims to find the transformation between the different views respective to the acquired point clouds (see Figure 2c) after the matching of key points. The initial alignment of two point clouds (typically called source  $Sr$  and target  $Tr$ , where  $Sr_{aligned} = T * Sr$ ) begins with the extraction of characteristic points; then the computation of the descriptors respective to these points, followed by the estimation of correspondences. After the rejection of bad matches, we estimate the transformation that aligns the sets of points and we apply it on the source point set. In our work, we initially adopted the matrix Singular Value Decomposition (SVD) [10] in order to estimate an initial guess of the underlying transform. This preliminary result is prone to error since it is computed fast without any sophisticated refinement (see Figure 3a). Nevertheless, it serves as a desirable initialization for the subsequent Iterative Closest Point (ICP) [10] alignment algorithm which gets frequently trapped in local minima.

#### 2) Iterative refinement with ICP

In most cases, the initial alignment does not give sufficiently accurate transformation. Hence, our goal now is to refine as much as possible the already obtained transformation with the ICP algorithm. The latter finds first the closest points in both clouds of points (correspondence estimation); then it proceeds through the estimation of a transformation that best aligns the matched correspondences (alignment). These two steps (correspondence, alignment) are repeated until reaching a termination criterion (typically, an error threshold or a maximum number of iteration). Figure 3b shows the outcome of applying the ICP algorithm on the SVD result. Although accurate when aligning a few point sets, ICP inherently leads to cumulative small errors due to a large number of views (around 50 views for a small room).

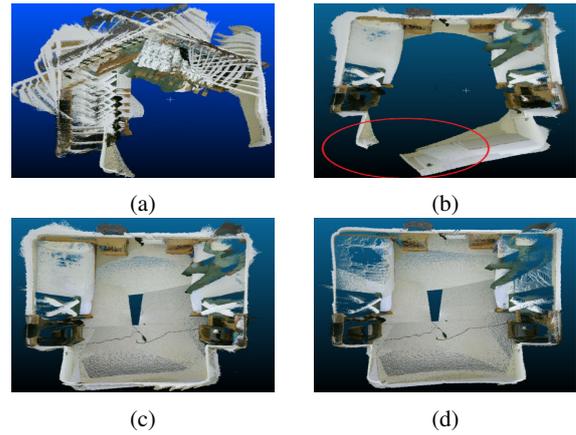


Fig. 3: Point cloud Alignment Results. (a) After initial alignment with SVD, (b) After refinement with ICP, (c) After loop closure, (d) After filtering and segmentation.

The global error between all the aligned 50 views after SVD is about 7m, which is huge for a  $4m \times 4m$  room. Indeed, ICP greatly minimizes this error to around 1m, but the accumulation of errors still impinges on the loop closure, due to a misalignment between the first and the last point set, that needs to be dealt with.

#### E. Loop Closure

As we mentioned earlier, point cloud alignment techniques introduce a non-negligible cumulative error, but the closure between the first and the last point clouds significantly reduced the error to its lowest levels. We noticed that this is not always the case, as when misalignment due to mis-correspondences is encountered, the overall reconstruction is no longer reliable. The reconstruction error increases with distance and misalignment is difficult to correct. Figure 4 illustrates the accumulation of misalignments after the application of SVD and ICP algorithms. This accumulation is more noticed at the corners of the room.

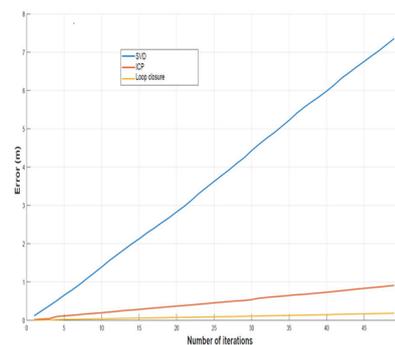


Fig. 4: The evolution of the cumulative error after SVD, ICP, and Loop Closure.

#### F. Filtering and segmentation

It is important that the point cloud after alignment bears little noise. For instance, the alignment algorithm filters some

**Input:**  $N$  point clouds  $C = \bigcup_{i=1}^N C_i$ .

**Output:** Optimal transformation ( $\hat{T}$ ) that minimizes the global misalignment ( $f$ ) between the  $C_i$ .

- 1) **Calculate the correspondences** between a given pair of clouds  $S, D$ ; such that  $(S, D) \in C \times C$  and  $S \neq D$ .
  - a) Initialize  $S' = \emptyset, D' = \emptyset$ .
  - b) For each point  $p \in S; \forall q \in D$ :
    - i) Calculate the distances between  $p$  and  $q$ .
    - ii) The point  $q$  of minimum distance corresponds to  $p$ .
    - iii) Remove  $p$  from  $S$ , and it to  $S'$ .
    - iv) Remove  $q$  from  $D$ , and it to  $D'$ .
  - c) Choose two subsets  $S''$  and  $D''$  of  $S'$  and  $D'$ , respectively, such that the elements of  $S''$  and  $D''$  are those of  $S'$  and  $D'$  with a distance below a given threshold.
- 2) **Estimate  $T$**  :

$$\hat{T} = \arg \min_T \left( \sum_{\forall S \neq D \in C} \|D - T * S\|^2 \right)$$

- 3) **Repeat 1 and 2** : until the following criterion is satisfied ( $f(\hat{T})$  is the global alignment error between all the  $N$  views after applying the transformation  $\hat{T}$ ) :

$$\|f(\hat{T})\|^2 \leq \epsilon$$

**Algorithm 1:** Loop closure Algorithm

of the noise but, if the cloud carries a lot of points not belonging to the model, the result of reconstruction may not be convincing to the investigator. After filtering our resulting point cloud with the Region Growing Segmentation filter [9] we obtained the result in Figure 3d.

### G. 3D triangulation and surface reconstruction

This step consists in creating facets from the point cloud by connecting the non-ordered points to each other in a triangular topology. To this end, we first apply Delaunay triangulation algorithm [9] for the creation of a smooth mesh. Figures 5a, 5b show our scene after surface reconstruction.

When we zoom in on the different parts of the scene (Figure 5b), we notice that surface reconstruction is generally of good quality except in some minor regions. These regions are typically characterized by a low density of 3D points resulting in a poor quality triangulation. To solve this problem, we captured 2D colored images during the acquisition phase, which will subsequently be used to fill the holes.

## IV. RESULTS AND DISCUSSION

In this section, we analyze the performance of our Loop Closure algorithm applied to improve the performance of 3D point cloud alignment. In order to demonstrate the intake of Loop Closure, we focus only on visual results and the variation of the cumulative error after the application of the three alignment algorithms. Figures 3c, 3d illustrate the final

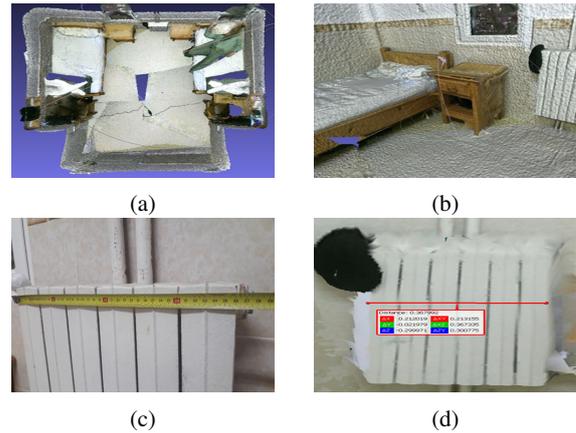


Fig. 5: Result After 3D Reconstruction. (a) Scene overview, (b) Zoom in the corner, (c) Real scene measurement, (d) 3D model measurement.

result of our scene after the alignment of all the views, we can see with the naked eye that the scene is closed, and the algorithm enhances the preceding result. We can also say that the alignment mean squared error has decreased considerably, it results in a misalignment of 18cm which is small given the 4m x 4m area of the room. The graph in Figure 4 illustrates this improvement.

### A. Implementation

The implementation of our proposed solution was achieved in C++ using the PCL library<sup>2</sup>, and Matlab 2017b on Windows 10 operating system. The different tests were carried out on a machine equipped with an Intel Core i5 processor of 2.30GHz and 6GB Ram.

In order to quantify the accuracy of our reconstruction, we compare the actual size and dimensions of scene objects to those obtained in the reconstructed model. This assessment is performed through cloud compare software<sup>3</sup>.

### B. Cloud Compare

This utility permits us the measurements of distances between 3D points. To make this comparison, we took real measurements of some objects in the crime scene, and we measured their respective reconstructed counterparts using Cloud Compare as shown in Figure 5c, 5d. Table I contains some measured dimensions of objects that were present in the crime scene. We notice clearly that our model gives enough accurate results, which were validated by investigators. These measurements mean that there is little overall difference between our model and the actual scene, where to 98% both measurements were similar. In addition, measurement error is more important in the larger objects due to the decreasing accuracy of the depth map delivered by the sensor.

<sup>2</sup><http://www.pointclouds.org>

<sup>3</sup><https://www.danielgm.net/cc/>

Objects	Measurements in centimeter (cm)		
	Real	3D Model	Error
Room width	384	400.4	16.4
Nightstand	46	46.6	0.6
Radiator	36	36.8	0.8
Chair	43	43.9	0.9
Cupboard door	74.5	75.7	1.2
Entrance door	99	99.2	0.2
Shelf	80.5	81	0.5
Table	85	86.1	1.1
Window	58	58.6	0.6
Bed width	97	97.4	0.4
Bed length	204	205.5	1.5
Electric socket	39	39.2	0.2

TABLE I: Comparison between actual measurements and model measurements.

C. Crime scenario

A good 3D reconstruction of the scene helps the investigator in the investigation, we simulated a crime scene that can enrich our work. This scenario is intended to reproduce the hypothesis prior to the action of committing the crime. To this end, we used Unity software <sup>4</sup>, which is a multi-platform game engine. This software allows us to introduce virtual objects into our scene and then to make animations and to create scenarios. In our case, we used two virtual persons (victim and criminal) as shows Figure 6. We programmed these characters in order to obtain a scenario as close as possible to reality.

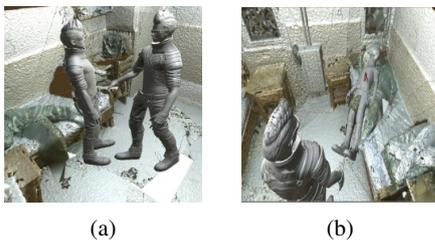


Fig. 6: Unity 3D simulated crime scenario: (a) Before, (b) After the crime.

D. Technical issues

We noticed during the deployment of our solution several technical problems. The first one concerns the stage of data acquisition, the Kinect sensor uses infrared rays to infer the depth of objects. Nevertheless, there are some hidden parts that could not be captured by the sensor. Another problem was encountered in the global alignment process, this problem is due to the accumulation of the errors. Correspondence estimation algorithms that are based on the feature descriptor are not optimized, which brings to bad matches, all of this leads to misalignment. The surface reconstruction stage generally gives good results, with the exception of the hidden or sparse regions. The last problem was noticed in Unity 3D during the simulation of crime scenarios, as the game engine accepts only three scene formats (.obj, .fbx, or .unity), we were obliged to

convert our model and to lose some scene’s color and geometry consistency.

V. CONCLUSION

We presented an RGB-D solution for accurate crime scene reconstruction and interactive scenario simulation. We leverage the potential of the cheap Kinect v2 TOF sensor in order to scan the crime space. Based on a set of key points we applied a preliminary alignment of the different views with SVD optimization. The latter is known for its speed but poor accuracy. However, it delivers a good guess for the subsequent iterative refinement algorithm (ICP). Afterward, we addressed the loop closure problem with a bundle adjustment algorithm.

Once all the views registered in the same reference frame, we proceeded through the triangulation and surface reconstruction. The resulting model is used for interactive crime investigation and object dynamics simulation. The obtained results show the effectiveness of our solution and its adequacy to the context being treated. We demonstrated the performance of our finding on a real scene. An accompanying video illustrates the whole chain of processing is available <sup>5</sup>.

As perspectives, we aim to extend the capabilities of the proposed solution to work outdoors. Regarding the external reconstruction performance of Kinect v2, we need to address the sensitivity to sunlight. Moreover, since our solution is better suited for offline processing, it would be interesting to investigate possible acceleration techniques for rapid on site reconstruction through the utilization of graphics processors and highly parallel techniques. Finally, the theory behind our work can be exploited for large scale reconstruction as well, after taking into account the required computational burden and the intervention of the different reconstruction elements.

REFERENCES

- [1] Simon Gibson and Toby Howard. “Interactive reconstruction of virtual environments from photographs, with application to scene-of-crime analysis”. In: *Proceedings of the ACM symposium on Virtual reality software and technology*. ACM. 2000, pp. 41–48.
- [2] Giovanna Sansoni, Marco Trebeschi, and Franco Docchio. “State-of-the-art and applications of 3D imaging sensors in industry, cultural heritage, medicine, and criminal investigation”. In: *Sensors* 9.1 (2009), pp. 568–601.
- [3] Erkan Bostanci. “3D reconstruction of crime scenes and design considerations for an interactive investigation tool”. In: *arXiv preprint arXiv:1512.03156* (2015).
- [4] Stephen Se and Piotr Jasiobedzki. “Instant scene modeler for crime scene reconstruction”. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)-Workshops*. IEEE. 2005, pp. 123–123.

<sup>5</sup><https://youtu.be/IYnJSNV7QkI>

<sup>4</sup><http://www.unity.org>

- [5] Trung Kien Dang, Marcel Worring, and The Duy Bui. "A semi-interactive panorama based 3D reconstruction framework for indoor scenes". In: *Computer vision and image understanding* 115.11 (2011), pp. 1516–1524.
- [6] Ursula Buck et al. "Accident or homicide–virtual crime scene reconstruction using 3D methods". In: *Forensic science international* 225.1-3 (2013), pp. 75–84.
- [7] Andreas Georgopoulos and Elisavet Konstantina Stathopoulou. "Data acquisition for 3D geometric recording: state of the art and recent innovations". In: *Heritage and Archaeology in the Digital Age*. Springer, 2017, pp. 1–26.
- [8] Jason Ligon et al. "3D point cloud processing using spin images for object detection". In: *2018 IEEE 8<sup>th</sup> Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 2018, pp. 731–736.
- [9] A Amamra. "Robust 3D registration and tracking with RGBD sensors". PhD thesis. Cranfield University, 2015.
- [10] Yilin Liu et al. "Curvature feature extraction based ICP points cloud registration method". In: *Optoelectronic Imaging and Multimedia Technology V*. Vol. 10817. International Society for Optics and Photonics, 2018, p. 1081707.

# 7<sup>th</sup> Workshop on Advances in Programming Languages

**P**ROGRAMMING 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
- Context-oriented programming languages to specify the behavior of software systems and dynamic adaptations
- 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
- Programming paradigms (aspect-oriented, functional, logic, object-oriented, etc.)
- Proof theory for programs
- Type systems
- Virtual machines and just-in-time compilation

- Visual programming languages

## STEERING COMMITTEE

- **Janousek, Jan**, Czech Technical University, Czech Republic
- **Luković, Ivan**, University of Novi Sad, Serbia
- **Mernik, Marjan**, University of Maribor, Slovenia
- **Slivnik, Boštjan**, University of Ljubljana, Slovenia

## EVENT CHAIRS

- **Varanda Pereira, Maria João**, Instituto Politecnico de Braganca, Portugal

## PROGRAM COMMITTEE

- **Barisic, Ankica**, Universidade Nova de Lisboa, Portugal
- **Fernandes, João Paulo**, Universidade de Coimbra
- **Horvath, Zoltan**, Eotvos Lorand University, Hungary
- **Janousek, Jan**, Czech Technical University, Czech Republic
- **Kardaş, Geylani**, Ege University International Computer Institute, Turkey
- **Kern, Heiko**, University of Leipzig, Germany
- **Kollár, Ján**, Technical University of Kosice, Slovakia
- **Kosar, Tomaž**, University of Maribor, Slovenia
- **Lopes Gançarski, Alda**, TELECOM SudParis, Evry, France
- **Luković, Ivan**, University of Novi Sad, Serbia
- **Mandreoli, Federica**, University of Modena, Italy
- **Martínez López, Pablo E. "Fidel"**, Universidad Nacional de Quilmes, Argentina
- **Mernik, Marjan**, University of Maribor, Slovenia
- **Milašinović, Boris**, University of Zagreb Faculty of Electrical Engineering and Computing, Croatia
- **Pai, Rekha**, National Institute of Technology Calicut, India
- **Papaspyrou, Nikolaos**, National Technical University of Athens, Greece
- **Porubán, Jaroslav**, Technical University of Kosice, Slovakia
- **Rangel Henriques, Pedro**, Universidade do Minho, Portugal
- **Saraiva, João**, Universidade do Minho, Portugal
- **Sierra Rodríguez, José Luis**, Universidad Complutense de Madrid, Spain
- **Slivnik, Boštjan**, University of Ljubljana, Slovenia



# Composition of Languages Embedded in Scala

Seyed H. HAERI (Hossein)

Université catholique de Louvain, Belgium

hossein.haeri@ucl.ac.be

Paul Keir

University of the West of Scotland, UK

paul.keir@uws.ac.uk

**Abstract**—Composition is amongst the major challenges faced in language engineering. Erdweg et al. offered a taxonomy for language composition. Mernik catalogued the use of the Language Definitional Framework LISA for composition sorts in that taxonomy. We produce a similar catalogue for embedded language engineering in Scala.

We begin with techniques that are not specific to Scala. They are applicable in any host language with a module system and support for higher order functions. We, then, present two more techniques to examine Scala-specific language engineering. Interestingly enough, even though dealing with embedded languages, in terms of lines of code, our material is of comparable length to its LISA counterpart. Our work lends insight into Scala’s serviceability for composition, as a host for embedded language engineering.

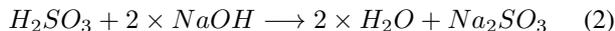
## I. INTRODUCTION

*a) Language composition is a piece of reality!:* Everyday, there are new programming languages that are born by combining ideas from older languages. Inspiration aside, that is an act of composition in many cases. For example, roughly put, Scala adds functional programming and ML modules with mixin composition to Java; which, in return, is C++ without pointers; which, in return, is C with OOP.

The taxonomy of Erdweg et al. [1] suggests a terminology and notations for describing such compositions. According to them, one can formalise our Scala description as:

$$\text{Scala} \cong \text{C} \triangleleft \text{C++} \triangleright \text{Java} \triangleleft (\text{MLModule} \uplus \text{Mixin}) \quad (1)$$

*b) Observations from Chemistry:* Consider the reaction:



In Chemistry, two key ingredients for success in the study of such equations are: (CI<sub>1</sub>) the availability of substances as the *subjects of study*, and, (CI<sub>2</sub>) knowledge about *how* to perform a desirable composition. In reaction (2), for instance, both substances  $H_2SO_3$  and  $NaOH$  need to be available. One also needs to know how to double  $NaOH$  for the equation balance to be right. Also, how to add  $NaOH$  to  $H_2SO_3$  (like the rate of addition, proper temperature, etc.) needs to be known.

*c) Programmatic Availability & Composition:* The study of formulae like equation (1) determines the precise relative position of languages. Using the outcome, one would be able to add, for example, what is missing in equation (1) so that the “ $\cong$ ” can be replaced by an “ $=$ ”. One would also gather that the left-out “FP $\uplus$ ” is necessary right before MLModule for the balance to be right. Such manipulations are similar

to adjusting coefficients in reaction (2) to obtain a balance. Similar to Chemistry, two key ingredients become noticeable here: (PLI<sub>1</sub>) *programmatic* availability of programming languages themselves and their belongings as the subjects of study, and, (PLI<sub>2</sub>) knowledge about how to *programmatically* obtain desirable language compositions.

By the time of this writing, (mainstream) languages are next to inaccessible as programmatic entities. The study of programmatic language composition, nonetheless, can be conducted independently using, say, contrived languages. That is how this paper tries to gain (PLI<sub>2</sub>).

*d) Contributions:* We demonstrate three techniques for composing languages embedded in Scala. The first (Section II) is applicable in any host language with a module system and support for higher order functions. The second (Section III) is based on Lightweight Modular Staging (LMS) [2]. And, the third – which is also a new solution to the Expression Problem (EP) [3], [4], [5] – employs (possibly restricted) abstract types. The trick in our third technique is promoting the cases of Algebraic Data Types (ADTs) into their own ADT-parameterised standalone components. We showcase each technique using the example compositions of Mernik [6]. We, then, compare the three techniques for their success in addressing the EP concerns (Section V). A discussion about the related work also comes at Section VI.

*e) Coding Conventions:* This paper assumes familiarity with Scala. For each showcase, the syntax and semantics come in separate packings called `syntax` and `semantics`, respectively. Due to space restrictions, in our code, the name of the showcase is only appended as a comment to the end of the first line of the respective `syntax` or `semantics`. For the same reason, our code is also otherwise unusually compressed. Whilst the showcases are referred to in the prose in CamlCase, their respective Scala package (containing the showcase’s `syntax` and `semantics`) is named `like_this` or abbreviated as `lt`.

## II. SCALA-UNSPECIFIC

Erdweg et al. catalogue five different ways languages can be composed: language extension, language restriction, language unification, self-extension, and extension composition. Mernik offers simple DSLs to showcase those ways in LISA [7]. In this section, we employ Mernik’s simple DSLs for the same purpose, albeit in Scala.

### A. Language Extension

A base language  $B$  is said to be extended to a language  $E$  when the description of  $B$  is amended with a description fragment to get  $E$ . Erdweg et al. denote that by  $B \triangleleft E$ . Consider the language `Robot` below (packaged under the name `robot` in Scala) for a robot arm that takes commands for moving one unit to either of the four 2D directions. The semantics of `Robot` involves updating the arm's position (recorded in terms of the  $x$  and  $y$  coordinates) based on the commands (lines 11 to 16).

```

1 object syntax { //robot
2   class Command
3   case object Left extends Command
4   case object Right extends Command
5   case object Up extends Command
6   case object Down extends Command
7   case class Commands(s: Seq[Command])
8   object semantics { import syntax._ //robot
9     class Position(var x: Int, var y: Int)
10    object position extends Position(0, 0)
11    def locate: Command => Unit = {
12      case Left => position.x -= 1
13      case Right => position.x += 1
14      case Up => position.y += 1
15      case Down => position.y -= 1
16    }
17    def locate(cs: Commands) = cs.s.foreach(locate)
18  }

```

`Robot` is extended to `RobotTime` (the `robot_time` package) by adding to the semantics, i.e.,  $\text{Robot} \triangleleft \text{RobotTime}$ :

```

1 package robot_time
2 import robot._; import syntax._
3 def time(cs: Commands): Int = cs.s.length

```

Assuming that executing each command takes one time unit, the total time required for a set of commands is the size of the set. The method `time` in line 3 above adds that piece of semantics to `Robot` to get `RobotTime`. whereas `Commands(Right, Down, Down)` in `Robot` has only got the semantics  $x = 1, y = -2$ , it also has the semantics  $t = 3$  in `RobotTime`. (The coordinates are obtained by `locate` in line 16 of `robot` and the timing by line 3 of `robot_time`.)

Here is a difference between our implementation of `RobotTime` and that of Mernik: The latter is done in LISA: a Language Definitional Framework (LDF) that combines OOP with Attribute Grammars (AGs) [8], [9]. As such, LISA's counterpart for `time` has to visit all the grammatical rules in `Robot` to attribute the new piece of semantics to them. On the contrary, Scala gave us the joy of simply equating `time` by the number of the commands, regardless of the grammatical rules involved.

### B. Language Restriction

A base language  $B$  is said to be restricted to a language  $R$  when certain parts of the  $B$ 's features are removed upon transition to  $R$ . This is denoted by  $B \triangleright R$ . A typical usage of that is when a language is narrowed to a core of it. That is, certain parts of the base syntax are cancelled into combinations of other base syntactic parts that are deemed to be equivalent. For example, both `GPH` [10] and `Utrecht HASKELL` [11] are developed like that.

The language `RobotPositive` below (packaged under `robot_positive`) restricts `Robot` to only `Up` and `Right` commands. (Technically, the object `syntax` below is not required. Yet, we retain it for completeness.)

```

1 object syntax { //robot_positive
2   import robot.syntax.{Right, Up, Commands}
3   object semantics { //robot_positive
4     import robot.syntax.{Right, Up, Commands}
5     import robot.semantics.position
6     def locate(cs: Commands) {
7       for(c <- cs.s) c match {
8         case Right => position.x += 1
9         case Up => position.y += 1
10      }
11    }
12  }

```

Any attempt to use the expression in the previous section under `RobotPositive` will fail to compile for the availability of `Down` in it, which is absent in `RobotPositive`. On the other hand, `Commands(Right, Up, Up)` has the semantics  $x = 1, y = 2$  under `RobotPositive`.

### C. Language Unification

Erdweg et al. say two languages  $L_1$  and  $L_2$  are unified to  $L$  when both  $L_1$  and  $L_2$  make sense independently from one another and from  $L$  (as the composition's outcome). Furthermore, in  $L$ , neither  $L_1$  nor  $L_2$  should be dominated by the other so that a concept of equity prevails in the composition. The notation is  $L = L_1 \uplus_g L_2$ , where  $g$  is the so-called glue code required for the composition.

Having seen the language `Robot`, we now consider the language `ExprAdd` (packaged under `expr_add`): a simple ADT with two cases for natural numbers and addition.

```

1 object syntax { //expr_add
2   class Expr { //Expr ::= Expr + Term | ...
3     def + (t: Term): Expr = Add(this, t)
4   }
5   class Term extends Expr { //Expr ::= ... | Term
6     //Term ::= n
7     case class Num(n: Int) extends Term
8
9     case class Add(left: Expr,
10                    right: Term) extends Expr
11  }
12  object semantics { //expr_add
13    import syntax._
14
15    def value: Expr => Int = {
16      case Num(n) => n
17      case Add(e, t) => value(e) + value(t)
18    }
19  }

```

Using `value` in line 15 above, one obtains the semantics 5, 12, and 6 for the expressions `Num(5)`, `Num(10) + Num(2)`, and `Num(1) + Num(2) + Num(3)`, respectively.

The language `RobotUniExprAdd` below (packaged under `robot_uni_expr_add`) unifies `Robot` and `ExprAdd` by allowing the robot arm to take commands for moving as many units to either of the four directions as the corresponding `ExprAdd` argument evaluates to. As such, `Commands(Right(Num(5)), Up(Num(2) + Num(10)), Up(Num(2) + Num(2) + Num(2)), Down(Num(4)))` has the semantics  $x = 5, y = 14$ . Check `locate` in line 17 below.

```

1 object syntax { //robot_uni_expr_add

```

```

2 import robot.syntax.Command
3 import expr_add.syntax._
4
5 case class Left(e: Expr) extends Command
6 case class Right(e: Expr) extends Command
7 case class Up(e: Expr) extends Command
8 case class Down(e: Expr) extends Command
9
10 object semantics { //robot_uni_expr_add
11 import robot.syntax.Commands
12 import robot_uni_expr_add.syntax._
13
14 import robot.semantics.position
15 import expr_add.semantics._
16
17 def locate(cs: Commands) {
18   for(c <- cs.s) c match {
19     case Left(e) => position.x -= value(e)
20     case Right(e) => position.x += value(e)
21     case Up(e) => position.y += value(e)
22     case Down(e) => position.y -= value(e)
23   }
24 }
25 }

```

#### D. Self Extension

This is the situation when the description of a language  $L$  itself is used for extending it. Typically, embedded DSLs self-extend their host language. For example, all the languages we present in this paper self-extend Scala.

Like Mernik, we believe that demonstrating self extension takes much more than the volume of a single research paper. This is because bootstrapping a language  $L$  to the level where it can handle self extension is already more involved than that volume. Hence, we too drop demonstration of self extension.

#### E. Extension Composition

Extension composition is when (both or at least one of) the language descriptions that are to be composed are themselves compositions of other language descriptions. As such, extension composition can be regarded as higher order composition. Six combinations of extension and unification are possible (three distinguished by Mernik):

- 1) Double-Unification ( $\uplus$ ):  $L_1 \uplus_g (L_2 \uplus_h L_3)$ .
- 2) Double-Extension ( $\triangleleft$ ):  $B \triangleleft E_1 \triangleleft E_2$ .
- 3) Extension by a Unification ( $\triangleleft(\uplus)$ ):  $B \triangleleft (L_1 \uplus L_2)$ .
- 4) Extension of a Unification ( $(\uplus)\triangleleft$ ):  $(L_1 \uplus L_2) \triangleleft E$ .
- 5) Unification with an Extension ( $\{\uplus, \triangleleft\}$ ):  $L \uplus (B \triangleleft E)$  or  $(B \triangleleft E) \uplus L$ . Note the symmetry.

We now consider each combination.

1) *Double-Unification* ( $\uplus$ ): To that end, we begin by presenting Mernik's language Dec (packaged under dec) in Scala. Dec enables the programmer to bind a set of variables to integer constants.

```

1 object syntax { //dec
2 case class ConstDefList(ds: Map[String, Int]) }

```

Unsurprisingly, the (Scala-automatic) semantics of ConstDefList("a" -> 5, "b" -> 10) is then  $\{a \mapsto 5, b \mapsto 10\}$ .

With that, we illustrate the first class of Mernik's extension compositions using RobotUniExprAddUniDec (packaged under rueaud). As suggested by its name, this language is (Robot $\uplus$ ExprAdd) $\uplus$ Dec. The Robot $\uplus$ ExprAdd portion is already presented. See robot\_uni\_expr\_add in Section II-C. We now show how to obtain the remaining unification.

```

1 import expr_add.syntax.{Expr, Term}
2 object syntax { //rueaud
3 import robot.syntax.Commands; import dec.syntax._
4 implicit class CDLInCs(val cdl:ConstDefList) {
5   def in (s: Commands) = {
6     consts = cdl.ds; new EnvComm(cdl.ds, s)
7   }
8 }
9 class EnvComm(val ds: Map[String, Int],
10               val cs: Commands)
11 var consts: Map[String, Int] = Map()
12 case class Var(n: String) extends Term
13 object semantics { //rueaud
14 import syntax._; import robot_uni_expr_add.syntax._
15 import robot.semantics._
16 def value_ext: (Expr, Expr => Int) => Int = {
17   case (Var(n), c) => consts(n)
18   case (e, c) =>
19     expr_add.ext_semantics.value_ext(e, c)
20 def value(e: Expr): Int = value_ext(e, value)
21 def locate(r: EnvComm) {
22   r.cs.s.foreach {
23     case Left(e) => position.x -= value(e)
24     case Right(e) => position.x += value(e)
25     case Up(e) => position.y += value(e)
26     case Down(e) => position.y -= value(e)
27   }
28 }

```

rueaud.syntax aims at reusing the former language descriptions as they are. To that end, it takes a *pimp my library* approach [12] on trying to implicitly (lines 4 to 10 above) give instances of dec.ConstDefList the extra feature of being followed by commands possibly referring to the declarations. Such declarations followed by expressions are then instances of EnvComm. The variable consts (line 11) is where the processed declarations are stored. The new ADT case Var (line 12) is for looking up the value a name is bound to. rueaud legitimises commands for moving the robot arm as many units as a pertaining expression evaluates to (lines 23 to 26). Note that, because of Var, those expressions can refer to declarations as well. All that together gives ConstDefList ("a" -> 5, "b" -> 10) in Commands(Right(Var("a")), Up(Num(2)+ Var("b")), Down(Num(4))) the semantics  $x = 5, y = 8$  in RobotUniExprAddUniDec.

Instead of reusing expr\_add.semantics.value, the rueaud.semantics.value method uses the method expr\_add.ext\_semantics.value\_ext, which will be explained shortly. This is because the former is closed on the set of ADT cases it can handle. Hence, we resort to the following *extensible* semantics of ExprAdd:

```

1 object ext_semantics {
2 import syntax._
3 def value_ext: (Expr, Expr => Int) => Int = {
4   case (Num(n), c) => n
5   case (Add(e, t), c) => c(e) + c(t)
6 def value(e: Expr): Int = value_ext(e, value)
7 }

```

In the fashion of  $\gamma\Phi C_0$  [13], value\_ext above takes a continuation argument c (line 3), which caters postponing the closing time until the appropriately complete shape [14] of the ADT is known (line 6 above for expr\_add and line 20 for rueaud). As such, extending RobotUniExprAdd to RobotUniExprAddUniDec here involves manipulating the former. See Section V for more.

2) *Double-Extension* ( $\triangleleft$ ): The idea in RobotTimeSpeed below (packaged under robot\_time\_speed) is to enable

the user to instruct the robot arm with the speed for its subsequent moves, until further notice. It adds a pertaining command to RobotTime to obtain Robot  $\triangleleft$  RobotTime  $\triangleleft$  RobotTimeSpeed.

```

1 object syntax { //robot_time_speed
2   import robot.syntax.Command
3   case class Speed(i: Int) extends Command
4 }
5 object semantics { //robot_time_speed
6   import syntax._; import robot.syntax.{Command, Commands}
7   import robot.semantics.position
8   def locate: Command => Unit = {
9     case Speed(_) => {}
10    case c => robot.semantics.locate(c)
11  }
12  def locate(cs: Commands) = cs.s.foreach(locate)
13  var speed: Double = 1.0
14  def time(cs: Commands): Double = {
15    var sum: Double = 0.0
16    for(c <- cs.s) c match {
17      case Speed(i) => speed = i
18      case _ => sum += (1.0 / speed)
19    }
20    sum
21  }

```

The new command for altering speed is Speed in line 3 above. This new command has no impact on the arm's position, as manifested in line 9. It is in the time calculation where, once used, the related variable (i.e., speed in line 12) is updated accordingly (line 16) and taken into consideration for subsequent commands (line 17). Commands(Up, Speed (2), Right, Left) has the semantics  $x = 1, y = 0, t = 2$  in RobotTimeSpeed.

3) *Extension by a Unification* ( $\triangleleft$  ( $\oplus$ )): We now demonstrate RobotExtExprAddUniDec = Robot  $\triangleleft$  (ExprAdd $\oplus$ Dec). We begin by ExprAddUniDec (packaged under eaud):

```

1 import expr_add.syntax._
2 object syntax { //eaud
3   import dec.syntax._
4   class EnvExpr(val ds: Map[String, Int], val e: Expr)
5   implicit class CDL2CDLInE(val cdl: ConstDefList) {
6     def in (e: Expr) = {
7       consts = cdl.ds
8       new EnvExpr(cdl.ds, e)
9     }
10  }
11  var consts: Map[String, Int] = Map()
12  case class Var(n: String) extends Term
13 }
14 object semantics { //eaud
15   import syntax._
16   import dec.syntax._
17   def value_ext: (Expr, Expr => Int) => Int = {
18     case (Var(n), c) => consts(n)
19     case (e, c) => expr_add.ext_semantics.value_ext(e, c)
20  }
21  def value(e: Expr): Int = value_ext(e, value)
22  def value(ee: EnvExpr): Int = value(ee.e)
23 }

```

eaud is similar to rueaud in Section II-E1 and we drop further explanation. RobotExtExprAddUniDec below (packaged under reeaud) tries to make use of eaud.

```

1 object syntax { //reeaud
2   import dec.syntax._; import robot.syntax.Commands
3   class EnvComm(val ds: Map[String, Int],
4                 val cs: Commands)
5   implicit class CDL2CDLInC(val cdl: ConstDefList) {
6     def in (s: Commands) = {
7       consts = cdl.ds
8       new EnvComm(cdl.ds, s)
9     }

```

```

10  }
11  var consts = eaud.syntax.consts
12 }
13 object semantics { //reeaud
14   import robot.semantics.position
15   import robot.uni_expr_add.syntax._
16   import eaud.semantics.value; import syntax._
17   def locate(r: EnvComm) {
18     r.cs.s.foreach {
19       case Left(e) => position.x -= value(e)
20       case Right(e) => position.x += value(e)
21       case Up(e) => position.y += value(e)
22       case Down(e) => position.y -= value(e)
23     }
24  }
25 }

```

Here are the few idiosyncrasies of reeaud: Firstly, reeaud fails to reuse most of the syntactic facilities of eaud. This is because the former employs declarations followed by commands, whereas the latter employs declarations followed by expressions. In line 11, nevertheless, consts is reused. Secondly, even though RobotExtExprAddUniDec = Robot  $\triangleleft$  ..., in reeaud.semantics, we do not reuse robot.syntax. On the contrary, in line 15, it reuses the syntax of robot\_uni\_expr\_add (for RobotUniExprAdd). This is because, in Robot, it is only possible to move the arm one unit to either direction. The Scala syntax for those two pieces of (embedded) syntax cannot coexist side by side. See Section III-A2 for more.

reeaud.semantics.locate is similar to rueaud.semantics.locate. In RobotExtExprAddUniDec, ConstDefList("a" -> 5, "b" -> 10)in Commands(Right(Var("a")), Up(Num(2)+ Var("b")), Down(Num(4))) has semantics  $x = 5, y = 8$ .

As pointed out by Mernik, so long as functionality is the only concern, RobotUniExprAddUniDec  $\equiv$  RobotExtExprAddUniDec. The difference, both in LISA and Scala, is in the language descriptions, and the combinations by which they are obtained. Unlike its LISA counterpart, nonetheless, obtaining RobotExtExprAddUniDec in Scala involves intermediate material that is not reused in the final product.

4) *Extension of a Unification* ( $\oplus$   $\triangleleft$ ): RobotUniExprAddExtRobotTime below (packaged under rueaert) extends RobotUniExprAdd (Section II) by a timing facility. The time required for carrying out a command of moving in one direction equals what the pertaining expression evaluates to (lines 9 to 12). The method time below is a simple fold operation on the given sequence of commands, based on that explanation. RobotUniExprAddExtRobotTime = (Robot  $\oplus$  ExprAdd)  $\triangleleft$  RobotTime.

```

1 object syntax { //rueaert
2   import robot_uni_expr_add.syntax._
3 }
4 object semantics { //rueaert
5   import expr_add.semantics._
6   import robot.syntax.Commands
7   import robot_uni_expr_add.syntax._
8   def time(cs: Commands): Int = (0 /: cs.s){
9     case (s, Left(e)) => s + value(e)
10    case (s, Right(e)) => s + value(e)
11    case (s, Up(e)) => s + value(e)
12    case (s, Down(e)) => s + value(e)
13  }
14 }

```

Commands (Right (Num(5)), Up (Num(2) + Num(10)),  
Up (Num(2) + Num(2) + Num(2)), Down (Num(4))) has  
the semantics  $x = 5, y = 14, t = 27$  in rueaert.

5) *Unification with an Extension* ( $\{\uplus, \triangleleft\}$ ):  
Take RobotUniExprMul = Robot  $\uplus$  ExprMul, where  
ExprAdd  $\triangleleft$  ExprMul. The language ExprMul extends  
ExprAdd by a new ADT case for multiplication (Mul).  
What is unique about ExprMul amongst the visited extension  
combinations is that, upon extension, it changes the syntactic  
categories of the ADT cases it borrows from ExprAdd.  
(And, in fact, it also provides a new syntactic category, i.e.,  
Factor.) As presented in Section III-B, this can impose  
a great deal of complexity when language extension is  
implemented using inheritance. Here is ExprMul (packaged  
under expr\_mul).

```

1 import expr_add.syntax.{Expr, Term, Add}
2 object syntax { //expr_mul
3   //Term ::= Factor | ...
4   class Factor extends Term
5   implicit class TermTimesFactor (val t: Term) {
6     def * (f: Factor): Term = Mul(t, f)
7   } //Term ::= ... | Term * Factor
8   //Factor ::= n
9   case class Num(n: Int) extends Factor
10  case class Mul(left: Term,
11               right: Factor) extends Term
12 }
13 object semantics { //expr_mul
14   import syntax._
15   import expr_add.
16     ext_semantics.{value_ext => add_value}
17
18   def value_ext: (Expr, Expr => Int) => Int = {
19     case (Num(n), c) => n
20     case (Add(e, t), c) =>
21       add_value(Add(e, t), c)
22     case (Mul(t, f), c) => c(t) * c(f)
23   }
24   def value(e: Expr): Int = value_ext(e, value)
25 }

```

In line 1, ExprMul imports the syntactic entities it borrows  
from ExprAdd: the ADT case Add and the syntactic categories  
Expr and Term. It then introduces its new syntactic category  
Factor in line 4. Next, in lines 5 to 7, it provides the syntactic  
sugar for multiplication. Note how it, afterwards, declares  
numbers to now be of the category Factor – as opposed to  
Term in expr\_add.syntax. The rest of expr\_mul should  
be straightforward except for the Scala syntax of lines 15  
to 16. Those lines abbreviate expr\_add.ext\_semantics  
.value\_ext to add\_value in expr\_mul.semantics. In  
line 21, expr\_mul reuses add\_value for the solo ADT case  
that it borrows from expr\_add, i.e., Add.

```

1 object syntax { //robot_uni_expr_mul
2   import robot.syntax.Command
3   import expr_add.syntax.Expr
4   import expr_mul.syntax._
5
6   case class Left (e: Expr) extends Command
7   case class Right (e: Expr) extends Command
8   case class Up (e: Expr) extends Command
9   case class Down (e: Expr) extends Command
10 }
11 object semantics { //robot_uni_expr_mul
12   import robot.syntax.Commands
13   import syntax._
14
15   import robot.semantics.position
16   import expr_mul.semantics._

```

```

17 def locate(cs: Commands) = cs.s.foreach {
18   case Left(e) => position.x -= value(e)
19   case Right(e) => position.x += value(e)
20   case Up(e) => position.y += value(e)
21   case Down(e) => position.y -= value(e)
22 }
23 }
24 }

```

The above implementation of RobotUniExprMul (packaged  
under robot\_uni\_expr\_mul) takes tightly after RobotUni-  
ExprAdd (in Section II-C). We, therefore, do not provide  
a dedicated walk-through. Commands (Right (Num(5) \* Num  
(2)), Down (Num(4) + Num(2) \* Num(3))) has the seman-  
tics  $x = 10, y = -10$  in robot\_uni\_expr\_mul.

### F. Language Specific?

To investigate the extent to which Scala-specific language  
features impact upon our design, we intend also to com-  
pare against realisations in other languages. To this end, we  
have prepared a C++ implementation which adopts the Scala  
approach outlined so far. Respecting the dynamic polymor-  
phism of the Scala original, the C++ implementation utilises  
shared\_ptr smart pointer to manage the memory allocation  
and runtime typing of expressions; allowing the vector  
container member object of the Commands class to store  
different expression types. User-defined integral and string  
literals also allow a notably concise syntax for the Num and Var  
instantiations; e.g., Commands{Right{"a"\_s}, Up{2\_n +  
"b"\_s}, Down{4\_n}}. Future work will explore this further.

Note that we are keen in the solution of this section not  
to employ Scala’s built-in open recursion. Due to unrelated  
reasons, however, Scala compilers might still employ open  
recursion internally to compile our code. Nonetheless, our  
code does not require that Scala idiosyncrasy. Testimony to  
that lack of requirement is our C++ code. Note that whilst  
open recursion is automatic in Scala, in C++, one needs to  
explicitly use “this->” for the late-binding of open recursion.

## III. LMS-BASED

Rompf and Odersky [2] coin Lightweight Modular Staging  
(LMS) for Polymorphic Embedding [15] of DSLs in Scala.  
They employ a fruitful combination of the Scala features  
detailed in [16] that, as a side-product, offers a very simple  
yet effective solution to EP. In this paper, we use LMS for  
that EP solution. The essence of LMS is the use of Scala  
traits for extensibility and super calls for reuse. With their  
mixin nature, Scala traits can extend one another, enjoying the  
benefits of inheritance. In particular, an ADT can be inherited  
upon trait extension. But, the heir trait can also add its own  
new ADT cases. On top of that, super calls enable reusing  
methods on the cases of the original ADT. Whereas the new  
cases can be handled by the same method, albeit overridden  
by the heir trait.

In the package eaud below (for ExprAddUniDec), for  
implementing both the syntax and semantics, traits are used  
– as opposed to objects in Section II. Instead of importing  
members from other languages, it now extends those other  
languages to acquire the same members via inheritance. In

Scala terms, `eaud.syntax` is, for instance, said to be mixing in `expr_add.syntax` and `dec.syntax`, in line 1 below.

In line 4, then, `eaud.semantics` overrides `value`. In line 5, it handles the new ADT case `eaud.syntax` introduces. All those other ADT cases that `eaud` inherits are, in line 6, relayed to the upper levels of inheritance.

```
1 trait syntax extends expr_add.syntax with dec.syntax {
2   ... /* like eaud.syntax in Section II-E3 */ ...}
3 trait semantics extends syntax with expr_add.semantics {
4   override def value: Expr => Int = {
5     case Var(n) => consts(n)
6     case e => super.value(e) } ...}
```

This is how LMS facilitates both simplicity and extensibility. (Note that we needed not to resort to `value_ext`.)

LMS has been successfully employed for languages in a multitude of applications. For the benefits of LMS, the reader is invited to consult those works. Given that we did not come to observe new benefits, we will not get into that here. We rather dedicate this section to the difficulties we faced over employing LMS for embedded language composition.

### A. Minor Difficulties

The two categories of minor difficulties we faced relate to language restriction (Section III-A1) and clashes occurred between names upon composition (Section III-A2).

1) *Language Restriction*: Upon extension, the programmer is usually provided with no means for acting selectively on the members to be inherited. When mixing traits too, all the (public or protected) members get inherited automatically. Hence, with inheritance being the means for language composition, language restriction is not possible. That enforces `import` as the fallback. With the use of traits, the mechanics is, however, more involved than Section II. Because traits are abstract, one needs to materialise them first (line 2 below), and only then, they can be `imported` from (line 3).

```
1 trait syntax/* robot_positive */{
2   val robosyn = new robot.syntax {}
3   import robosyn.{Right, Up, Command, Commands}
4 }
```

Even though LISA also employs inheritance for language composition, this difficulty does not arise there. The reason is as follows: Being also an AG system, (subject) language semantics is specified in LISA by traversing the concrete syntax. On the other hand, leveraging its OOP, LISA allows the heir language to override the parent language's concrete syntax. As a result, language restriction is also possible in LISA via inheritance.

One final related comment: In our experience, enforced `imports` like those required for language restriction were not exclusive to that way of language composition. In fact, in a good number of other occasions, the languages do make selective use of one another. That, on its own, was not a knotty problem. It, however, requires increasingly more care when it comes to interplay with hierarchies of languages and the relevant Scala mixins.

Note that `imported` names (like those in line 3 above) do not get inherited but the respective materialised traits

(like `robosyn` in line 2 above) do. Such `imports` can be required on several occasions down the hierarchy. In the case of unification, however, where the multiple inheritance nature of mixins is employed, an extra `override` might also be enforced to disambiguate duplicated names across the meeting two hierarchies. See Section III-B for more.

2) *Name Clash*: Recall from Section II-E3 that `RobotExtExprAddUniDec = Robot < (ExprAdd ⊕ Dec)`. In an LMS-based implementation of `RobotExtExprAddUniDec`, therefore, one would naturally want to implement `rueaud.semantics` as follows:

```
1 trait semantics extends rueaud.syntax with
2   robot.semantics with eaud.semantics {//rueaud
3   ... /* locate like Section II-E1 */ ...}
```

That is, however, not possible. The error message is: “object `Left` is not a case class, nor does it have an `unapply` /`unapplySeq` member.” The problem is that, even though `Left` is inherited from `robot`, in `locate`, Scala would not be able to match it using the syntax `Left(e)`. The available constructor and extractor of `Left` take no arguments. Moreover, overloading that syntax is not possible. This is because Scala desugars both case classes and case objects to objects with `unapply` (or `unapplySeq`) methods. Objects, on the other hand, are final, banning any later manipulation. To proceed, one needs to use `robot_uni_expr_add.semantics` in return of `robot.semantics`.

The problem is harder to diagnose for `RobotUniExprAddExtRobotTime`. Recall from Section II-E4 that `RobotUniExprAddExtRobotTime = (Robot ⊕ ExprAdd) < RobotTime`. For the attempt

```
1 trait semantics extends rueaert.syntax with
2   robot_uni_expr_add.semantics with
3   robot_time.semantics {... /* rueaert */ ...}
```

even when one employs `robot_uni_expr_add.semantics` instead of `robot.semantics`, one gets an error – this time, regarding the composition itself: “overriding object `Left` in trait `syntax`; object `Left` in trait `syntax` cannot override final member.” The problem here is with `robot_time` being an extension to `robot`, bringing the case object `Left` into the mix with that of `robot_uni_expr_add` that takes an argument.

### B. Major Difficulties

The difficulties we spoke about in the previous subsection were not particularly acute in that not many circumvention attempts would fail for them. In this section, we will report a multi-staged combat with an acute difficulty we faced. In short, the combat was against the combination of Scala's path-dependant typing and intervention of concrete syntax.

The contents of this section might look too specific to Scala. They are not. Scala's path-dependant typing is just one way to foster family polymorphism [17] (as opposed to lightweight family polymorphism [18]). The familiar reader will figure out that the same problem is likely to emerge in every host language that embraces family polymorphism.

Given that `ExprMul` is a direct extension to `ExprAdd`, one's first guess would be:

```
1 trait expr_mul.syntax extends expr_add.syntax {...}
```

That is, however, not possible because, then, Num cannot be overridden. Recall from Section II-E5 that ExprMul changes the syntactic category of Num. But, even an attempt like those in Section III-A1 for the syntax

```
1 trait syntax {
2   val easyn = new expr_add.syntax {} //expr_mul
3   import easyn.{Expr, Term, Add} /* Num, Factor, etc. */
4 }
```

would still cause failure for the semantics.

```
1 trait expr_mul.semantics extends syntax with
2   expr_add.semantics {...}
```

Here is the error message: “overriding object Num in trait syntax; object Num in trait syntax cannot override final member.” This is because of the clash between the Num of such a `expr_mul.syntax` and `expr_add.semantics`. See Section III-A2 for an explanation on similar error messages.

Now, let us suppose for the sake of argument that the semantics too selectively `imports` the ADT cases:

```
1 trait semantics { //expr_mul
2   val emsyn = new expr_mul.syntax {}
3   import emsyn.{Num, Mul, Factor}
4   val easyn = new expr_add.syntax {}
5   import easyn.{Expr, Add, Term}
6   ... /* value or value_ext here */ ...
7 }
```

Recall that ExprMul adds the ADT case Mul to ExprAdd. To reuse – à la LMS – the ExprAdd semantics whilst also handling the new ADT case, one may (mistakenly) try:

```
1 override def value: Expr => Int = {
2   case Mul(t, f) => value(t) * value(f) ...
3 }
```

But, that will not type-check because of path-dependant typing interference: Expr in value’s signature is different from Expr that Mul inherits from. Here is the error message for line 2 above: “constructor cannot be instantiated to expected type; found: semantics.this.emsyn.Mul required: semantics.this.Expr.” Even worse: An attempt for reusing the semantics of the only ADT case that remains intact over the move from ExprAdd to ExprMul using `value_ext`

```
1 trait semantics {... //expr_mul
2   import easem.{value_ext => add_value}
3   def value_ext: (Expr, Expr => Int) => Int = {
4     case (Num(n), c) => n
5     case (Add(e, t), c) => add_value(Add(e, t), c)
6     case (Mul(t, f), c) => c(t) * c(f) }
```

will again fail due to path-dependant typing. The error message for line 5 above is: “type mismatch; found: semantics.this.easyn.Add required: semantics.this.easem.Expr.”

Given that `expr_mul.semantics` is to reuse pattern matching of `expr_add.semantics`, the former is also bound to the types – here, ADT cases – of the latter. In order to prevent the path-dependant clashes, thus, the only way forward seems to be for **both** `expr_mul.syntax` and `expr_mul.semantics` to import types of `expr_add.semantics`. This is, of course, very unnatural for the former.

```
value: Expr => Int
expr_add{Num, Add}, eaud{Num, Add, Var},
      expr_mul{Num, Add, Mul}
locate: Command => Unit (without e)
robot{Right, Left, Up, Down},
      robot_positive{Right, Up}
locate (with (e))
in reeaud: EnvComm => Unit
in robot_uni_expr_add: Commands => Unit
in rueaud: EnvComm => Unit
in robot_uni_expr_mul: Commands => Unit
EnvComm
reeaud, rueaud
```

Fig. 1: Duplicate Entities in Sections II and III

```
1 trait syntax { //expr_mul
2   val easem = new expr_add.semantics {}
3   import easem.{Expr, Term, Add}; ...
4   trait semantics extends syntax { //expr_mul
5     import easem.{Expr, Add, value_ext => add_value}
6     ...
7     def value_ext: (Expr, Expr => Int) => Int = {
8       case (Num(n), c) => n
9       case (a: Add, c) => add_value(a, c)
10      ...
11    }
12    ...
13 }
```

Still, if not done craftily enough, path-dependant typing can be an impediment. Replacing the line 9 above with

```
case (a @ Add(_, _), c) => add_value(a, c)
```

will fail to type-check because `a` is considered to be of type `this.Add`; whereas, `add_value` accepts an `easem.Expr`. The unsightly circumvention would be:

```
case (a @ Add(_, _), c) => add_value(a,
asInstanceOf[easem.Expr], c.asInstanceOf[easem
.Expr => Int]).
```

We would like to remind that all the difficulties illustrated in this section were only experienced in the presence of manipulation in the syntactic categories upon extension. Syntactic categories are often used for dealing with concrete syntax. Semantics, on the other hand, inputs abstract syntax. The following section presents a solution that disassociates concrete syntax from abstract syntax. It applies the LMS at the abstract syntax level, and, hence, independently of the concrete syntax that varies across languages. That design sets the different languages free on engineering their syntactic categorisation whilst enjoying the benefits of LMS.

#### IV. REFACTORING

The previous two sections were developed as if the guest language implementer was not aware in advance of the next guest languages and the upcoming combinations. We also maintained a backward compatibility policy in that we did not touch the older languages as we proceeded. Refactoring, however, is common in everyday software development.

Refactoring can have a variety of meanings, depending on the target and the methods used [19]. Here, we do not plan extensive refactoring. We only focus on duplicate elimination in the fashion of the *extract superclass* method [19, §12.6]. Fig. 1 lists a number of duplicates in Sections II–III.

We notice that the method `value` is duplicate in `expr_add`, `eaud`, and `expr_mul`. More precisely, the ADT cases `Num` and `Add` – which are, basically, inherited from `expr_add` – are handled thrice in the codebase. As will be shown in this section, we gave `value` its own abstraction.

We also notice that the method `locate` is present in two sets of language descriptions: in (i) `robot` and `robot_positive` (when the four direction commands do not take arguments); and, in (ii) `reeaud`, `robot_uni_expr_add`, `rueaud`, and `robot_uni_expr_mul` (when the four direction commands do take arguments). Each of those sets constitutes a candidate for refactoring. Finally, `EnvComm` is common between `reeaud` and `rueaert` – constituting yet another refactoring candidate. Although we have indeed refactored the candidates of this paragraph as well, we will not include their demonstration in this paper. The interested reader can look them up in our online codebase.

Let us now focus on refactoring the first row of Fig. 1. (Refactoring the other rows of Fig. 1 is done similarly.) Here is a succinct summary of actions to be taken: The idea is a combination of LMS and Component-Based Mechanisation [20], [21], [13]. We parameterise the ADT cases `Num`, `Add`, `Var`, and `Mul` by the language description and perform their semantics evaluation independently of the language description. We pack the two former cases – namely, `Num` and `Add` that are common between all the items in the first row of Fig. 1 – together in a trait. Then, we extend that trait for `Var` and later for `Mul`, both *à la* LMS. Finally, the concrete language descriptions only get to mix the respective abstract descriptions. The elaboration follows.

```

1 trait na_syntax {
2   type E
3   type N <: E //N for Num
4   type A <: E //A for Add
5
6   def n_extr(n: N): Option[Int]
7   def a_extr(a: A): Option[(E, E)]
8
9   object N {def unapply(n: N) = n_extr(n)}
10  object A {def unapply(a: A) = a_extr(a)}
11 }
12 trait na_semantics extends na_syntax {
13   def value: E => Int = {
14     case N(n) => n
15     case A(e1, e2) => value(e1) + value(e2)
16   }
17 }

```

In the trait `na_syntax` above, the abstract type `E` (in line 2) is a language-independent representation for the expression type of a guest language. Such a guest language can be an item in row 1 of Fig. 1 or any similar language with integer arithmetics that at least contains integral literals and addition. Given that ADTs are implemented in Scala using plain inheritance, two more language-independent abstract types have been employed that are announced to be extending `E`. Those are `N` for `Num` and `A` for `Add`, in lines 3 and 4.

Because `N` and `A` are supposed to later be instantiated to the respective cases of an ADT, they are expected to come with the Scala matching syntax, like those in lines 14 and 15. The Scala machinery for enforcing availability of the desirable matching syntax requires a discipline in coding that

is slightly tricky. The discipline involves, for each ADT case abstract type, inclusion of a same-named (singleton) object – called *companion* object – that ships, then, with an *extractor*, i.e., an *unapply* method of the right type signature. The actual duty of the extractor is relayed to an abstract method, to be enforced to every guest language that implements `na_syntax`. For `N`, for instance, that duty is on `n_extr` in line 6. The Scala signature of `n_extr` means that, if matching `N` succeeds, it would be initialising an argument of type `Int`. All that wiring enables the method `na_semantics.value` to handle the semantics of `Num` and `Add`.

```

1 trait nam_syntax extends na_syntax {
2   type M <: E
3   def m_extr(m: M): Option[(E, E)]
4   object M {def unapply(m: M) = m_extr(m)}
5 }
6 trait nam_semantics extends nam_syntax with na_semantics {
7   override def value: E => Int = {
8     case M(e1, e2) => value(e1) * value(e2)
9     case e => super.value(e)
10  }
11 }

```

The trait `nam_syntax` adds the abstract type `M` (in line 2 above), which corresponds to `Mul`. It also provides the Scala matching syntax in lines 3 and 4. The trait `nam_semantics` reuses (*à la* LMS) what is already implemented by `na_semantics` by performing a *super* call on the relevant ADT cases (line 9).

```

1 trait expr_add.syntax extends na_syntax {
2   /* ... like lines 2 to 10 of
3     expr_add.syntax in Section II ... */
4   type E = Expr //Fix the ADT type.
5   type N = Num //Fix the Num case.
6   type A = Add //Fix the Add case.
7   //And, fix the extractors.
8   def n_extr(n: Num) = Num.unapply(n)
9   def a_extr(a: Add) = Add.unapply(a)
10 }
11 trait expr_add.semantics extends
12   expr_add.syntax with na_semantics

```

In addition to working out the Section II concrete syntax, the trait `expr_add.syntax` above, now is required to provide evidence on it indeed having ADT cases for integral literals and addition. That, again involves some slightly tricky discipline consisting of two steps. First, in lines 4 to 6, the concrete counterparts for the abstract (ADT case) types in `na_syntax` are fixed. Second, in lines 8 and 9 the extractors promised to `na_syntax` are fixed.

Recall from `expr_add.syntax` of Section II that `Num` and `Add` are both case classes. Scala actually desugars case classes to normal classes in addition to companion objects with the right-typed *unapply* methods. That is why we can use `Num.unapply` and `Add.unapply` off-the-shelf.

Nothing more remains for `expr_add.semantics` to do except inheriting its (abstract and concrete) syntax from `expr_add.syntax` and its semantics from `na_semantics`.

```

1 trait expr_mul.syntax extends nam_syntax {
2   val easyn = new expr_add.syntax {}
3   import easyn.{Expr, Term, Add};...
4   //like lines 4 to 11 of expr_mul.syntax in Section II...
5   type E = Expr; type N = Num; type A = Add; type M = Mul
6   def n_extr(n: Num) = Num.unapply(n)
7   def a_extr(a: Add) = Add.unapply(a)
8   def m_extr(m: Mul) = Mul.unapply(m)

```

```

9 trait ExprMulSemantics extends
10   ExprMulSyntax with NamSemantics

```

Implementing `ExprMul`, in this fashion, is similar, as demonstrated above. It only is that, like in Section III, our use of traits instead of objects in favour of LMS imposes instantiation of the trait `ExprAddSyntax` (line 2) before `importing` the desirable concrete syntax items (line 3).

#### Remarks

`na_semantics` is similar to how one defines the semantics of `Num` and `Int` using Modular Structural Operational Semantics (MSOS) [22]. In MSOS, the semantics of a component is defined exclusively in terms of the relevant language elements – making it ignorant about all other language elements. `na_semantics` only concerns `Num` and `Int`, and, is ignorant about other language elements.  $\gamma\Phi C_0$  [13] describes that as: “client  $na\_semantics \langle F \triangleleft Int \oplus Num \rangle \{ \dots \}$ ,” where  $F$  is the *family parameter* of  $na\_semantics$ . In words, that reads: A family  $\Phi$  to be substituted for  $F$  needs at least to have components *Int* and *Num* (or their equivalents) in its mix.

From another language theoretical viewpoint, `na_syntax` and `na_semantics` are both type classes [23]. From that viewpoint, `ExprAddSyntax` is an instance of `na_syntax` and `ExprAddSemantics` is an instance of `na_semantics`. The evidence for the former is provided in lines 2 to 7 in `na_syntax`. Interestingly, however, our encoding of type classes in Scala is not the common one [24]. In particular, we do not prescribe the use of `implicit`s.

As also announced at the last paragraph of Section III, `na_syntax` and `na_semantics` (and also `nam_syntax` and `nam_semantics`) relate to the abstract syntax only. This is how they leverage LMS and yet do not suffer from the concrete syntactic anomalies discussed in Section III. Moreover, unlike Modular Reifiable Matching [25], the technique we presented in this section is not exclusively targeting two-level types [26]. The reason is that our technique in this section fully disassociates concrete syntax from the abstract syntax so there no longer is an issue of levels in the types. LMS itself comes with no such separation either – suggesting the name *abstract LMS* for our technique.

It is noteworthy that the disassociation of abstract and concrete syntax with the lack of the LMS anomalies discussed in Section III needs not specifically be *à la* LMS. The same impact can also be achieved using integration of a decentralised pattern matching [27]. In the latter technique, the syntax is defined in terms of abstract syntax components. The concrete syntax in the latter technique is then defined on top of those syntax components. The difference is that the abstract LMS composes components (that correspond to ADT cases) *additively* [28, §17.3], whilst the latter technique would be composing them *sequentially*.

The connection between this technique and Component-Based Software Engineering (CBSE) [28, §17],[29, §10] is also interesting. From a CBSE standpoint, `nam_syntax` is a component in that: Without binding to a particular implementation, it specifies its so-called ‘requires’ and ‘provides’

interfaces. The `nam_syntax` ‘requires’ interface is its lines 2 and 3 – imposing the following two requirements, respectively: The user of `nam_syntax` needs to provide a type `M`. And, there has to be a way to extract two expressions of type `E` from an instance of `M`. In return, the ‘provides’ interface of `M` is its line 4, where `M`’s Scala match syntax (used in line 8 of `nam_semantics`) is offered. As such, `nam_syntax` is promoting the ADT case `Mul` to its standalone component.<sup>1</sup> This is an important characteristic of the third technique that relates to the EP. Next section is dedicated to that relationship.

## V. EXPRESSION PROBLEM

EP is a recurrent problem in the field of Programming Languages, for which a wide range of solutions have thus far been proposed, e.g., [31], [32], [33]. Consider [34], [35], [36], [31], [32], [33], to name a few. Haeri [21] defines EP as the challenge of finding an implementation for an ADT – defined by its cases and the functions on it – that:

- E1. is *extensible in both dimensions*, i.e., both new cases and functions can be added.
- E2. provides *weak static type safety*, i.e., applying a function  $f$  on a statically<sup>2</sup> constructed ADT term  $t$  should fail to compile when  $f$  does not cover all the cases in  $t$ .
- E3. upon extension, forces *no manipulation or duplication* to the existing code.
- E4. accommodates *separate compilation*, i.e., compiling the extension imposes no requirement for repeating compilation or type checking of existing code. Such static checks should not be deferred to the link or run time.

In Sections II–IV, we presented three techniques for embedded language composition in Scala. All the three techniques satisfy E4. We now reflect on their E1–E3 competence: The first technique clearly satisfies E1. Section III-A2 outlines a scenario where LMS fails to satisfy E1. Whether the third technique satisfies E1 depends on whether it employs trait mixing for composition or not. Note that it needs not. The three techniques all relax E2, although they can be circumvented to work when defaults are available [35]. That is a consequence of Scala performing pattern matching at runtime. LMS too relaxes E2 and that has thus far been considered an acceptable setting. (For example, MVCs [37] and Torgersen’s second solution [34] both have the same issue.) The state of affairs for LMS might change in future though [38].

As witnessed by `RobotUniExprAddUniDec` in Section II-E1, the Scala-unspecific technique fails to satisfy E3 when new cases are to be added. As detailed in Section III-B, LMS has to fight path-dependant typing to satisfy E3 when syntactic categories are updated upon composition. Whether there always is a winning strategy for LMS in such a situation is not known. The third technique clearly satisfies E3.

<sup>1</sup>Two reasons for not promoting `Num` and `Add` to components: 1) that would complicate presentation. 2) the current design in which those two ADT cases are packed together in a single component (i.e., `na_syntax`) demonstrates how to address the Common Reuse Principle of Martin [30].

<sup>2</sup>If the guarantee was for dynamically constructed terms too, we would have called it strong static type safety.

We understand that the path-dependant typing difficulties of the LMS-based technique might indeed be a result of our peculiar design. In particular, our choice of giving the syntax and semantics of a language each a trait of their own might be picked as the root cause. We would like to defend that choice of ours, specifically, for the likelihood of engineering (or experimentation with) more than one semantics for the same syntax [15]. In such cases, separation of the syntax and semantics is inevitable.

Finally, one may wonder whether the third technique makes it to a new solution to EP. The answer is indeed yes. At least for EP in presence of defaults [35]. This is the third EP solution of its kind: It promotes ADT cases to their own ADT-parameterised components. See [20], [21] for the first and [27] for the second EP solution of this kind.

## VI. RELATED WORK

*a) LISA:* As stated earlier, this paper is highly inspired by Mernik [6]. We essentially took his examples for showing how to compose languages embedded in Scala. With LISA being an LDF, even though Scala is famous for its hospitality to embedded languages, we were surprised to end up having less lines-of-code (LoC) in all the three techniques.

Fig. 2 summarises the LoC comparison. In the LoC there, we have also included some syntactic cosmetics that we did not display in this paper. In our experience, the occasions where Scala outperforms LISA by far are those where the task was a ready cake for GPLs. Examples are RobotTime for all the techniques and RobotExtExprAddUniDec for the third technique. For the former, a simple container size query does the job. For the latter, simple trait mixing does.

The first technique generally performs better (in terms of LoC) than LISA. The second is even better usually with its utilisation of trait mixing (dismissing the obvious `imports` and `super` calls. At last, the third is the best with its a posteriori refactoring. The two occasions when LISA considerably outperforms Scala are RobotUniExprAdd for the first technique and RobotUniExprMul for the third. Those correspond to Sections III-A2 and III-B, respectively.

The factored out code in the third technique is not counted in Fig. 2. Once that too is added, the total LoC reaches 328 – which is 2 more than first technique’s LoC. We tend to think the reason is the simplicity in the semantics of Mernik’s examples. That caused the number of lines the refactoring saves to be less than the extra overhead the technique requires. For more realistic case studies, we expect the balance to be completely different. That would be well in favour of refactoring due to reasonably more involved semantics.

*b) Other Language Composition Catalogues:* Völter [39] proposes a taxonomy of language composition that he showcases in JetBrains MPS. His taxonomy is along axes, not all of which having a clear correspondent in the work of Erdweg et al. As explained by Mernik, the resulting ways for language composition that Völter prescribes, however, are subsumed by the latter taxonomy. Völter’s taxonomy gives (syntax-oriented) IDE development for languages a higher weight.

Barrett, Bolz, and Tratt [40] catalogue composition of six different Python and Prolog virtual machines. Their study has a particular focus on measuring performance of the resulting interpreters upon composition.

Zhang et al. [41] facilitate composition of languages that are embedded using Object Algebras [42]. This is achieved using their simple predesignated annotation. Their showcase focuses on hierarchies of language extension. Using linearised multiple language inheritance, they also simulate a single language unification. Zhang et al. do not consider higher order composition.

Melange [43] is an LDF that is specially equipped for language composition. Various syntactic facilities are available in Melange to instruct mix-and-match for many different aspects of a language – ranging from syntax, dynamic and static semantics, and name-binding to IDE features. Language composition under Melange is catalogued for a small set of showcases but with in-length discussions on customisability. The current documentation of Melange, however, makes it hard for us to compare its catalogue of language composition with similar works. Specifically, we fail to figure out which ways for language composition Melange supports in general (namely, for other scenarios than the ones already in their documentation) and how.

*c) Components for Language Specification:* PPlanCompS funcons are syntactic constructs that ship with their own fixed static and dynamic semantics (presented in MSOS). The PPlanCompS specification of a programming language is developed by merely assembling funcons [44]. Example assemblies are larger academic languages [45] and medium-scale ones [46]. Despite their merit, funcons do not constitute CBSE components. In particular, funcons do not ship with their ‘requires’ interfaces.

MVCs [37] are components for solving an extension to EP. Rather than components in their CBSE sense, however, MVCs are components in a Component-Oriented Programming [47] sense. (Cf. [21, §4.3].) MVCs rely on the implementation details of `how` a component realises its interfaces. CBSE components, in contrast, are identified by their ‘requires’ and ‘provides’ interfaces.

Haeri and Schupp [20], [27] take a CBSE approach for the implementation of embedded languages. Their approach employs type constraints and multiple inheritance. The third technique here employs (possibly constrained) abstract types instead of type parameters. Although essentially the same, the former can make code terser. In Scala, however, offering the match syntax is apparently not possible for type parameters.

Finally, Cazzola and Vacchi [48] too have taken a CBSE approach. Their components correspond to a DSL’s compiler passes. Accordingly, how their work relates to the common language specification formalisms is not clear. In contrast, components in our third technique are ADT cases – acting as the unit of study for formal semantics.

*d) Component-Based AGs:* AGs are a powerful means for language specification with many benefits that are well-studied. Attempts to modularise AGs go back to Saraiva and

	$L_1$	$L_2$	$L_3$	$L_4$	$L_5$	$L_6$	$L_7$	$L_8$	$L_9$	$L_{10}$	$L_{11}$	$L_{12}$	$L_{13}$	Sum
LISA	42	23	13	19	19	32	39	41	20	34	23	20	19	344
$T_1$	32	7	16	26	34	11	40	25	31	34	17	22	31	326
$T_2$	30	6	15	20	29	10	34	20	26	28	13	23	33	287
$T_3$	29	5	15	16	16	10	10	20	23	6	13	23	16	202

Columns:  $L_1$  = Robot,  $L_2$  = RobotTime,  $L_3$  = RobotPositive,  $L_4$  = ExprAdd,  $L_5$  = RobotUniExprAdd,  $L_6$  = Dec,  $L_7$  = RobotUniExprAddUniDec,  $L_8$  = RobotTimeSpeed,  $L_9$  = ExprAddUniDec,  $L_{10}$  = RobotExtExprAddUniDec,  $L_{11}$  = RobotUniExprAddExtRobotTime,  $L_{12}$  = ExprMul,  $L_{13}$  = RobotUniExprMul Rows: LISA = Mernik’s Implementation,  $T_i$  = Technique  $i$ , for  $i \in \{1, 2, 3\}$

Fig. 2: Lines-of-Code Comparison between Mernik’s LISA and Our Three Techniques

Swierstra [49]. Saraiva’s Higher Order AGs (HOAGs) [50] were the initial steps towards using AGs in a component-based fashion. Viera and Swierstra [51] formally define several ways to combine HOAGs. However, those ways do not tightly correspond to the usual composition mechanics of general-purpose languages.

So long as EP is concerned, the correct behaviour of a HOAG w.r.t. E2 is not universally agreed upon. In terms of HOAGs, that amounts to the absence of an attribute expected from another component in the mix. In particular, should the code then fail statically or dynamically? Zipper functions [52], [53] act like HASKELL by statically reporting such errors so long as they can be caught iteratively [54].

Kiama [55] uses AGs embedded in Scala for language specification. It is possible to use Kiama in a component-based fashion – as done for embedding Oberon-0 [56] in Scala [57]. However, disassociation of the concrete and abstract syntax can become non-trivial in Kiama. We anticipate that would cause similar difficulties to those we faced over our second technique. For the Oberon-0 embedding, facing such difficulties were unlikely for the different pieces of syntax were all available in advance. On the contrary, whilst composing unrelated pieces of syntax, clash of concrete syntax is likely.

## VII. CONCLUSIONS AND FUTURE WORK

In this paper we present three different techniques for composing languages embedded in Scala. The first is Scala-unspecific and works in presence of common module systems and higher order functions (Section II). The second is LMS-based and requires mixin composition and `super` calls (Section III). The third works by promoting ADT cases to ADT-parameterised components (Section IV). We showcase the three techniques using the example compositions of Mernik, which, in return, were designed to exhibit LISA’s composition facilities for Erdweg et al.’s taxonomy of composition. We manifest the strengths and weaknesses of each technique. We compare them according to their performance as EP solutions (Section V) and LoC (Section VI-0a).

Systematic study of embedded language composition is a young topic. Numerous paths exist for future research. Examining our third technique against larger testcases is an immediate future work. A promising candidate is the LDTA’11 challenge of modular implementation of Oberon-0. The testcase can then be compared with the LDTA’11 contestants. We anticipate complications in dealing with a few

issues along the way: Firstly, the technique takes a design-by-contract approach on the names it chooses for abstract types, e.g., `A` and `N` in `na_syntax`. In large scale, these names are likely to clash upon composition. Avoiding that would imply a priori knowledge. That kind of knowledge is, however, rare in experimental language design. Secondly, outside lab settings, usual software engineering techniques may become inevitable. We took the lab liberty of not being concerned with that here. For example, `position` and `consts` lack proper scoping and are common intact amongst all the descendants of `Robot` and `Dec`, respectively.

Type classes are more widely practised in HASKELL. It would be interesting to see our third technique in HASKELL with its type classes instead of Scala’s mixins and inheritance. The comparison between the results of ours and those according to the following two HASKELL EP solutions would be particularly interesting: Data Types a la Carte [36] and Parametric Compositional Datatypes [32].

Object Algebras are gaining gravity as a powerful abstraction for embedded language development [31], [58], [59], [41]. The current technology for embedding Object Algebras, however, is heavyweight in both term creation [60] and algebra composition. It is easy to turn `na_syntax` and the like into Object Algebra Interfaces to lower those two weights. How useful the result would be in lowering those two weights in the current Object Algebras technology is another future work.

Finally, it is important to also produce catalogues like this paper in other host languages than Scala. Many languages have merits in hosting other languages. But, the limits of that and the key factors of it are not clear. Composition of the embedded languages is certainly amongst the important factors. A head-to-head comparison on hospitality of language composition is missing. We are currently working on that.

## REFERENCES

- [1] S. Erdweg, P. G. Giarrusso, and T. Rendel, “Language Composition Untangled,” in 12<sup>th</sup> LDTA, A. Sloane and S. Andova, Eds. ACM, Mar. 2012, p. 7.
- [2] T. Rompf and M. Odersky, “Lightweight Modular Staging: a Pragmatic Approach to Runtime Code Generation and Compiled DSLs,” in 9<sup>th</sup> GPCE. Eindhoven, Holland: ACM, 2010, pp. 127–136.
- [3] W. R. Cook, “Object-Oriented Programming Versus Abstract Data Types,” in FOOL, ser. LNCS, J. W. de Bakker, W. P. de Roever, and G. Rozenberg, Eds., vol. 489, Holland, Jun. 1990, pp. 151–178.
- [4] J. C. Reynolds, “User-Defined Types and Procedural Data Structures as Complementary Approaches to Type Abstraction,” in *New Direc. Algo. Lang.*, S. A. Schuman, Ed. INRIA, 1975, pp. 157–168.
- [5] P. Wadler, “The Expression Problem,” Nov. 1998, Java Genericity Mailing List.

- [6] M. Mernik, "An Object-Oriented Approach to Language Compositions for Software Language Engineering," *J. Sys. & Soft.*, vol. 86, no. 9, pp. 2451–2464, 2013.
- [7] M. Mernik, M. Lenic, E. Avdicausevic, and V. Zumer, "LISA: An Interactive Environment for Programming Language Development," in *11<sup>th</sup> CC*, ser. LNCS, R. N. Horspool, Ed., vol. 2304. Springer, Apr. 2002, pp. 1–4.
- [8] D. E. Knuth, "Semantics of Context-Free Languages," *Math. Sys. Theo.*, vol. 2, no. 2, pp. 127–145, 1968.
- [9] J. Paakki, "Attribute Grammar Paradigms - A High-Level Methodology in Language Implementation," *ACM Comp. Surv.*, vol. 27, no. 2, pp. 196–255, 1995.
- [10] P. Trinder, K. Hammond, H.-W. Loidl, and S. Peyton Jones, "Algorithm + Strategy = Parallelism," *JFP*, vol. 8, no. 1, pp. 23–60, Jan. 1998.
- [11] A. Dijkstra, J. Fokker, and S. D. Swierstra, "The Architecture of the Utrecht HASKELL Compiler," in *2<sup>nd</sup> HASKELL*, S. Weirich, Ed. Edinburgh, Scotland: ACM, 2009, pp. 93–104.
- [12] M. Odersky, "Pimp my Library," *Artima Developer Blog*, vol. 9, Oct. 2006.
- [13] S. H. Haeri and S. Schupp, "Expression Compatibility Problem," in *7<sup>th</sup> SCSS*, ser. EPiC Comp., J. H. Davenport and F. Ghourabi, Eds., vol. 39. EasyChair, Mar. 2016, pp. 55–67.
- [14] J. Jeuring, S. Leather, J. P. Magalhães, and A. R. Yakushev, "Libraries for Generic Programming in HASKELL," in *Adv. Func. Prog., 6<sup>th</sup> Int. School, AFP*, ser. LNCS, P. W. M. Koopman, R. Plasmeijer, and S. D. Swierstra, Eds., vol. 5832. Springer, May 2008, pp. 165–229.
- [15] C. Hofer, K. Ostermann, T. Rendel, and A. Moors, "Polymorphic Embedding of DSLs," in *7<sup>th</sup> GPCE*, Y. Smaragdakis and J. G. Siek, Eds. Nashville, TN, USA: ACM, Oct. 2008, pp. 137–148.
- [16] M. Odersky and M. Zenger, "Scalable Component Abstractions," in *20<sup>th</sup> OOPSLA*. San Diego, CA, USA: ACM, 2005, pp. 41–57.
- [17] E. Ernst, "Family Polymorphism," in *15<sup>th</sup> ECOOP*, ser. LNCS, J. Lindskov Knudsen, Ed., vol. 2072. Springer, Jun. 2001, pp. 303–326.
- [18] C. Saito, A. Igarashi, and M. Viroli, "Lightweight Family Polymorphism," *J. Func. Prog.*, vol. 18, no. 3, pp. 285–331, 2008.
- [19] M. Fowler, "Refactoring: Improving the Design of Existing Code," in *2<sup>nd</sup> XP/Agile*, ser. LNCS, D. Wells and L. A. Williams, Eds., vol. 2418. Springer, Aug. 2002, p. 256.
- [20] S. H. Haeri and S. Schupp, "Reusable Components for Lightweight Mechanisation of Programming Languages," in *12<sup>th</sup> SC*, ser. LNCS, W. Binder, E. Bodden, and W. Löwe, Eds., vol. 8088. Springer, Jun. 2013, pp. 1–16.
- [21] S. H. Haeri, "Component-Based Mechanisation of Programming Languages in Embedded Settings," Ph.D. dissertation, STS, TUHH, Germany, Dec. 2014.
- [22] P. D. Mosses, "Modular Structural Operational Semantics," *JLAP*, vol. 60–61, pp. 195–228, 2004.
- [23] P. Wadler and S. Blott, "How to Make ad-hoc Polymorphism Less ad-hoc," in *16<sup>th</sup> POPL*. ACM Press, Jan. 1989, pp. 60–76.
- [24] B. C. d. S. Oliveira, A. Moors, and M. Odersky, "Type Classes as Objects and Implicits," in *25<sup>th</sup> OOPSLA*, W. R. Cook, S. Clarke, and M. C. Rinard, Eds. ACM, Oct. 2010, pp. 341–360.
- [25] B. C. d. S. Oliveira, S.-C. Mu, and S.-H. You, "Modular Reifiable Matching: A List-of-Functors Approach to Two-Level Types," in *8<sup>th</sup> HASKELL*, B. Lippmeier, Ed. ACM, Sep. 2015, pp. 82–93.
- [26] T. Sheard and E. Pasalic, "Two-Level Types and Parameterized Modules," *JFP*, vol. 14, no. 5, pp. 547–587, 2004.
- [27] S. H. Haeri and S. Schupp, "Integration of a Decentralised Pattern Matching: Venue for a New Paradigm Inter-marriage," in *8<sup>th</sup> SCSS*, ser. EPiC Comp., M. Mosbah and M. Rusinowitch, Eds., vol. 45. EasyChair, Apr. 2017, pp. 16–28.
- [28] I. Sommerville, *Software Engineering*, 9<sup>th</sup> ed. Addison-Wesley, 2011.
- [29] R. S. Pressman, *Software Engineering: A Practitioner's Approach*, 7<sup>th</sup> ed. McGraw-Hill, 2009.
- [30] R. C. Martin, "Design Principles and Design Patterns," 2000, online article available from the [ObjectMentor](#) website.
- [31] B. C. d. S. Oliveira and W. R. Cook, "Extensibility for the Masses – Practical Extensibility with Object Algebras," in *26<sup>th</sup> ECOOP*, ser. LNCS, vol. 7313. Springer, 2012, pp. 2–27.
- [32] P. Bahr and T. Hvitved, "Parametric Compositional Data Types," in *4<sup>th</sup> MSFP*, ser. ENTCS, J. Chapman and P. B. Levy, Eds., vol. 76, Feb. 2012, pp. 3–24.
- [33] Y. Wang and B. C. d. S. Oliveira, "The Expression Problem, Trivially!" in *15<sup>th</sup> Modularity*. New York, NY, USA: ACM, 2016, pp. 37–41.
- [34] M. Torgersen, "The Expression Problem Revisited," in *18<sup>th</sup> ECOOP*, ser. LNCS, M. Odersky, Ed., vol. 3086, Oslo (Norway), Jun. 2004, pp. 123–143.
- [35] M. Odersky and M. Zenger, "Independently Extensible Solutions to the Expression Problem," in *FOOL*, Jan. 2005.
- [36] W. Swierstra, "Data Types à la Carte," *JFP*, vol. 18, no. 4, pp. 423–436, 2008.
- [37] B. C. d. S. Oliveira, "Modular Visitor Components," in *23<sup>rd</sup> ECOOP*, ser. LNCS, vol. 5653. Springer, 2009, pp. 269–293.
- [38] T. Rompf, "Reflections on LMS: Exploring Front-End Alternatives," in *7<sup>th</sup> SIGPLAN Symp. Scala*, A. Biboudis, M. Jonnalagedda, S. Stucki, and V. Ureche, Eds. ACM, Nov. 2016, pp. 41–50.
- [39] M. Völter, "Language and IDE Modularization and Composition with MPS," *GTTSE*, vol. 7680, pp. 383–430, 2011.
- [40] E. Barrett, C. F. Bolz, and L. Tratt, "Approaches to Interpreter Composition," *Comp. Lang., Sys. & Struct.*, vol. 44, pp. 199–217, 2015.
- [41] H. Zhang, Z. Chu, B. C. d. S. Oliveira, and T. van der Storm, "Scrap Your Boilerplate with Object Algebras," in *29<sup>th</sup> OOPSLA*, J. Aldrich and P. Eugster, Eds., Oct. 2015, pp. 127–146.
- [42] Gutttag, J. V. and Horning, J. J., "The Algebraic Specification of Abstract Data Types," *Acta Informatica*, vol. 10, pp. 27–52, 1978.
- [43] T. Degueule, B. Combemale, A. Blouin, O. Barais, and J.-M. Jézéquel, "Melange: A Meta-Language for Modular and Reusable Development of DSLs," in *8<sup>th</sup> SLE*, R. F. Paige, D. Di Ruscio, and M. Völter, Eds., Oct. 2015, pp. 25–36.
- [44] P. D. Mosses, "Component-Based Description of Programming Languages," in *BCS Int. Acad. Conf.*, E. Gelenbe, S. Abramsky, and V. Sassone, Eds. Brit. Comp. Soc., 2008, pp. 275–286.
- [45] P. D. Mosses and F. Vesely, "FunKons: Component-Based Semantics in  $\mathbb{K}$ ," in *WRLA*, ser. LNCS, S. Escobar, Ed., vol. 8663. Springer, Apr. 2014.
- [46] M. Churchill, P. D. Mosses, N. Sculthorpe, and P. Torrini, "Reusable Components of Semantic Specifications," *Trans. Aspect-Orient. Soft. Dev. XII*, vol. 12, pp. 132–179, 2015.
- [47] M. D. McIlroy, "Mass Produced Software Components," in *Proc. NATO Conf. Soft. Eng.* New York, US: Petrocelli/Charter, 1969, pp. 138–155.
- [48] W. Cazzola and E. Vacchi, "Language Components for Modular DSLs using Traits," *ComLan*, vol. 45, pp. 16 – 34, 2016.
- [49] J. Saraiva and D. Swierstra, "Generic Attribute Grammars," in *2<sup>nd</sup> WAGA*, vol. 99, 1999, pp. 185–204.
- [50] J. Saraiva, "Component-Based Programming for Higher-Order Attribute Grammars," in *1<sup>st</sup> GPCE*, ser. LNCS, D. S. Batory, C. Consel, and W. Taha, Eds., vol. 2487. Springer, Oct. 2002, pp. 268–282.
- [51] M. Viera and D. Swierstra, "Attribute Grammar Macros," *Sci. Comp. Prog.*, vol. 96, pp. 211–229, 2014.
- [52] P. Martins, J. P. Fernandes, J. Saraiva, E. Van Wyk, and A. Sloane, "Embedding Attribute Grammars and their Extensions using Functional Zippers," *Sci. Comp. Prog.*, vol. 132, pp. 2–28, 2016.
- [53] J. P. Fernandes, P. Martins, A. Pardo, J. Saraiva, and M. Viera, "Memoized Zipper-Based Attribute Grammars and their Higher Order Extension," *Sci. Comp. Prog.*, vol. 173, pp. 71–94, 2019.
- [54] A. Middelkoop, A. Dijkstra, and D. Swierstra, "Iterative Type Inference with Attribute Arammars," in *9<sup>th</sup> GPCE*, E. Visser and J. J., Eds. ACM, Oct. 2010, pp. 43–52.
- [55] A. M. Sloane, "Lightweight Language Processing in Kiama," in *GTTSE III*, ser. LNCS, J. M. Fernandes, R. Lämmel, J. Visser, and J. Saraiva, Eds., vol. 6491. Springer, Jul. 2009, pp. 408–425.
- [56] N. Wirth, *Compiler Construction*, ser. Int. Comp. Sci. Series. Addison-Wesley, 1996.
- [57] A. M. Sloane and M. Roberts, "Oberon-0 in Kiama," *Sci. Comp. Prog.*, vol. 114, pp. 20–32, 2015.
- [58] B. C. d. S. Oliveira, T. van der Storm, A. Loh, and W. R. Cook, "Feature-Oriented Programming with Object Algebras," in *27<sup>th</sup> ECOOP*, ser. LNCS, G. Castagna, Ed., vol. 7920. Montpellier, France: Springer, 2013, pp. 27–51.
- [59] T. Rendel, J. I. Brachthäuser, and K. Ostermann, "From Object Algebras to Attribute Grammars," in *28<sup>th</sup> OOPSLA*, A. P. Black and T. D. Millstein, Eds. ACM, Oct. 2014, pp. 377–395.
- [60] A. P. Black, "The Expression Problem, Gracefully," in *MASPEGHI@ECOOP 2015*, M. Sakkinen, Ed. ACM, Jul. 2015, pp. 1–7.

# Supporting Source Code Annotations with Metadata-Aware Development Environment

Ján Juhár

Department of Computers and Informatics  
Technical University of Košice  
Letná 9, 042 00 Košice, Slovakia  
Email: jan.juhar@tuke.sk

**Abstract**—To augment source code with high-level metadata with the intent to facilitate program comprehension, a programmer can use annotations. There are several types of annotations: either those put directly in the code or external ones. Each type comes with a unique workflow and inherent limitations. In this paper, we present a tool providing uniform annotation process, which also adds custom metadata-awareness for an industrial IDE. We also report an experiment in which we sought whether the created annotating support helps programmers to annotate code with comments faster and more consistently. The experiment showed that with the tool the annotating consistency was significantly higher but also that the increase in annotating speed was not statistically significant.

## I. INTRODUCTION

THE MAIN hindrance programmers deal with when they need to comprehend source code is known as the *abstraction gap*. This gap exists between the problem domain and the solution domain of a given software system. Many high-level concerns from the problem domain are either lost or scattered as programmers transform them to code. As argued by LaToza *et al.* [1] and Vranić *et al.* [2], programmers often ask questions about the *intent* behind particular source code fragments. In this paper, we present and evaluate an approach for helping to preserve the high-level knowledge within source code annotations.

### A. Motivation

Two general approaches for retrieving information otherwise lost or scattered in source code are available:

- *recovery* of pieces of high-level information from the code by means of reverse engineering, and
- *preservation* of a programmer's thoughts and intentions in software artifacts.

Feature location tools from the recovering approach usually produce list of source code elements that are evaluated as relevant to the given feature [3]. Preserving approaches directly assign high-level information to source code elements with annotations [4], [5]. Although in both cases the retrievable data represent *source code metadata* (abbreviated: *metadata*), they are of different nature.

This work was supported by project VEGA No. 1/0762/19: "Interactive pattern-driven language development" and grant of FEEI TUKE no. FEI-2018-55: "Methods of code classification based on knowledge profiles".

Recovering approaches use *intrinsic metadata*, which either define source code elements themselves or can be derived from these elements. In contrast to them, preserving approaches focus on *extrinsic metadata*, which complement the intrinsic ones by adding custom, high-level details explicitly recorded by programmers. On one side, the more accurate preserved knowledge may help to bridge the abstraction gap better than the lower-level recovered one. On the other side, recording programmer's mental model of the code brings in an additional cost: the programmer must spend extra time to record it.

The immediate availability of intrinsic metadata makes them a great choice for code analysing tools in integrated development environments (IDEs) [6]. These can provide *structure-aware* visualizations (e.g., file structure browsers, semantic code highlighting, linting) and actions (e.g., contextual code completion, refactoring). However, intrinsic metadata also restrict these tools to lower-level domains.

It is thus a worthwhile question whether adding the metadata upfront will be too costly compared to any benefits they may bring later. Sulír *et al.* show in [4] how concern metadata in the form of Java annotations can enable rapid construction of reader's mental model of the implementation. Report of Ji *et al.* [7] shows that presence of feature-related metadata within source code comments was beneficial for software product line development. The authors presume that by employing a supporting tool the benefits can be further increased.

To tackle the tool support for such custom metadata in the source code we need to consider both the type of annotations that can be used and granularity of elements where they can be used. Ideally, a programmer would not have to consider all the different ways in which metadata can be bound to the code, but directly *express the intention* to bind metadata to specific source code elements and let a tool to perform the binding. This is the main motivational factor for the work presented in this paper, proposed through the idea of uniform annotation process in a metadata-aware development environment.

### B. Goal

The goal of this paper is twofold. First, we present the idea and prototype implementation of the uniform annotation process in an integrated development environment (IDE) extended by metadata-awareness. By IDE metadata-awareness we mean the ability to work with both custom code-bound metadata and

with annotations that bind them as with first-class source code elements. This goal is addressed in Section II.

Second, in sections III and IV we report an experiment we performed to evaluate the effect our prototype tool has on annotating speed and consistency of comment annotations created during code annotating task. An overarching research question for the experiment is “*Does metadata-aware IDE help programmers to annotate code with comment annotations?*”.

## II. ANNOTATION PROCESS IN A METADATA-AWARE DEVELOPMENT ENVIRONMENT

IDE tools are adapted to use intrinsic metadata derived from code elements. They can easily bind them to the originating elements and build dynamic code views or projections from them [6], [8]. Extrinsic metadata are available also but mostly limited to data from version control and bug tracking systems. As such, they are bound only to files, or lines of text. We can achieve more specific bindings with source code annotations.

In our work the term *source code annotation* (abbreviated: *annotation*) has a more general meaning than, e.g., *Java annotation*. As per Definition 1, we consider any binding of metadata to source code element as annotation.

**Definition 1.** *Annotations are in-place or addressing bindings of custom metadata to source code elements.*

A development environment able to utilize the metadata recorded by code authors may provide program comprehension support on a higher level of abstraction, closer to the problem domain of a software system. Our idea of such *metadata-aware development environment* (MADE) comprises of three following aspects:

- 1) Support for the annotation process, during which a programmer binds metadata to code elements.
- 2) Preservation of annotations and metadata as code changes.
- 3) Utilization of the metadata in various IDE tools to facilitate program comprehension.

In the work presented in this paper, we focus on the first aspect: on supporting the *annotation process*, which we define in Definition 2.

**Definition 2.** *Annotation process, or annotating, is a process in which metadata are being bound to code elements.*

### A. Types of Source Code Annotations

When faced with a task to annotate code, a programmer has three following types of annotations to chose from:

- Internal annotations contained within the source code files, further classifiable into two distinct types:
  - *Language-level annotations* (LLAs), which use native programming language constructs for metadata.
  - *Structured comment annotations* (SCAs), which give the “metadata” status to code comments.
- *External annotations* (EAs), which are created with a supporting tool and bound to the source code by addressing the annotated elements.

Each of these annotation types has a different set of inherent limitations, which we discuss in the following.

1) *Language-level Annotations*: LLAs are formally defined in language’s grammar and all standard language tools can work with them. On the other side, they can be used only if the language itself does support them, and only on elements where it supports them. An example of applying custom metadata with LLAs in *Java* language is given in Listing 1.

Listing 1. *Java* annotations as high-level metadata

```
@NoteChange @TagManagement
public void addTag(String tag) { /* ... */ }
```

2) *Structured Comment Annotations*: SCAs reuse general code comments, which can contain arbitrary text. For that reason we need to define a specific syntax for them that would allow us to parse the metadata. Listing 2 shows an example of such syntax. Such annotations can be used in almost any language, considering that a comment can be put at the desired place in the code. But they require supporting tool that can recognize the metadata and bind them to specific code elements.

Listing 2. High-level metadata in structured comment annotation

```
// [# note change ] [# tag management ]
public void addTag(String tag) { /* ... */ }
```

3) *External Annotations*: EAs are superimposed over the code by means of an addressing mechanism that locates annotated elements. The mechanism can use simple addresses like element’s starting and ending offsets within a file, or more robust descriptors of code elements [9]. Annotations are usually visualized in the code editor (see Fig. 1).

The most significant advantage of EAs is that arbitrary code fragments can be annotated, even inside files the programmer cannot (or does not want to) modify. On the other side, their addresses need to be kept in sync with changes made to the code and they are completely dependent on a supporting tool.

### B. Supporting the Annotation Process

Our focus on custom extrinsic metadata allows us to assume that annotating is going to be performed “manually” by programmers. Their goal may be to capture their mental model of the code in a form that can be used by tools and can help future maintainers of the code. In the design of a tool



Fig. 1. Metadata bound to code fragment through external annotations displayed in the editor’s gutter

supporting such annotation process, we should strive to remove unnecessary distractions from programmers' primary goal of annotating. For us, it primarily means that regardless of which type of annotation is used the workflow should be the same.

However, annotation of a given type differs from other types in how exactly it is applied to the code and what conditions must be met before it can be applied. The most important differences are the following.

- *Definition*: LLAs are represented by language elements, which may need to be defined before they can be applied (e.g., like *Java*'s annotation types). Similarly, EAs may require definition through a tool [10]. SCAs have no single definition of the source metadata and the programmer needs to maintain them individually in each comment.
- *Application*: Internal annotations must be typed<sup>1</sup> into the code at the appropriate place. EAs are applied through environment's UI and their application may be preceded by code selection.
- *Binding*: LLAs are bound to specific elements in-place according to language's grammar. EAs use addressing bindings, which may be text-level or element-level. And comments, as free-standing statements, have no bindings to the surrounding code elements.

To deal with these differences, we designed an *abstracted annotation process*, which by itself does not require any changes to the annotated code and imitates annotating the code with EAs. In this process, operations required to annotate code fragments<sup>2</sup> should be performed through IDE actions. The actions cover selecting code fragment for annotation and selecting annotation representing required metadata. When these are selected, the annotation should be applied automatically with the configured annotation type.

We implemented a prototype tool called *Connotator*<sup>3</sup>, which supports the abstracted annotation process and its configuration per project. The tool is implemented as a plug-in for *JetBrains IntelliJ* platform-based IDEs. The current implementation supports all three annotation types for the *Java* language, and SCAs and EAs for languages *Kotlin* and *Python*. It also provides *source code editor augmentations* [11] related to code annotations. In the following, we describe the annotation process and its realization in *Connotator* in more detail.

1) *Defining metadata annotations*: All annotations representing custom metadata are in *Connotator* managed through the main tool window (see Fig. 2). Annotations defined there may be applied to the code. The metadata model is currently rather simple: annotations are defined only by their names and optionally they can have a parent annotation. The tool supports annotation name refactoring, which appropriately updates all their existing instances in the code.

2) *Selecting annotatable code fragment*: Only valid code selections—those matching some AST elements—are mean-

<sup>1</sup>LLAs may take advantage of already existing IDE support like code completion, but some typing is still involved.

<sup>2</sup>We use the term code fragment in a sense of one or more consecutive elements selected for annotation.

<sup>3</sup>*Connotator* is available at <https://git.kpi.fe.i.tuke.sk/jan.juhar/connotator>.

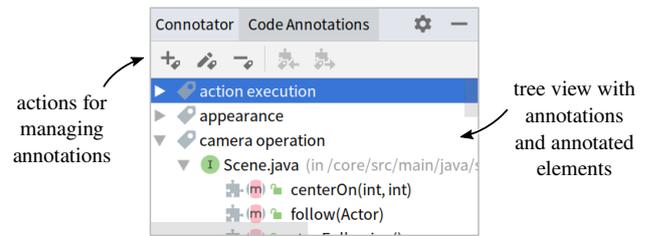


Fig. 2. Annotations tool window.

ingful targets for metadata. In general, fine-grained element selection (like statements and expressions) should be possible, but the specific set of annotatable elements should be configurable per-project. These requirements are in *Connotator* met through code fragment selection facility using tree patterns matched against the PSI tree<sup>4</sup> of code elements. A user can specify these patterns as *XPath*-like expressions built from a set of basic element types. For example, the following expression can be used to match *Java* statements without block bodies:

```
codeBlock/statement[not child::blockStatement]
```

To select elements for annotation, the user uses fragment-selecting action (with keyboard shortcut or from menus). The action resolves annotatable elements from current text selection or caret position in the code editor, as can be seen in Fig. 3(a).

3) *Applying annotations*: Existing metadata annotation can be applied on selected code fragment with dedicated action that allows the user to specify the annotation. Its usage is shown in Fig. 3(b). Once the annotation is selected, *Connotator* finishes code annotation automatically. It selects the annotation type to use from the tool's configuration (it can be specified separately for each type of annotatable elements) or selects one automatically based on their availability and predefined priority (LLA > SCA > EA). Fig. 3(c) shows the result of applying an annotation on class fields when *Java* LLAs are configured for their element type. Note that the tool also generates required *Java* annotation types if they do not exist yet.

The same annotation process is applicable for any annotation type; the only difference is in the final alteration of the code. As an example, Fig. 4 shows a `for` statement annotated with SCA. All the source code editor augmentations, like gutter icon and annotation highlighting, remain the same. An EA would differ only by no visible annotation inserted into the code.

### C. Binding Comments to Code Elements

Going back to the differences among annotation types, the one we did not discuss so far is *binding*. As far as LLAs and EAs are concerned, the binding is defined either by the language's grammar or by specific addressing mechanism used by the tool supporting EAs<sup>5</sup>. On the other hand, SCAs do not have any grammar-based or other bindings to surrounding code

<sup>4</sup>PSI (Program Structure Interface) tree is a version of concrete syntax tree backing most structure-aware features of the *IntelliJ* platform [12].

<sup>5</sup>In *Connotator*, we currently use just a very basic offset-based addressing. A more robust solution is out of scope of here presented work.

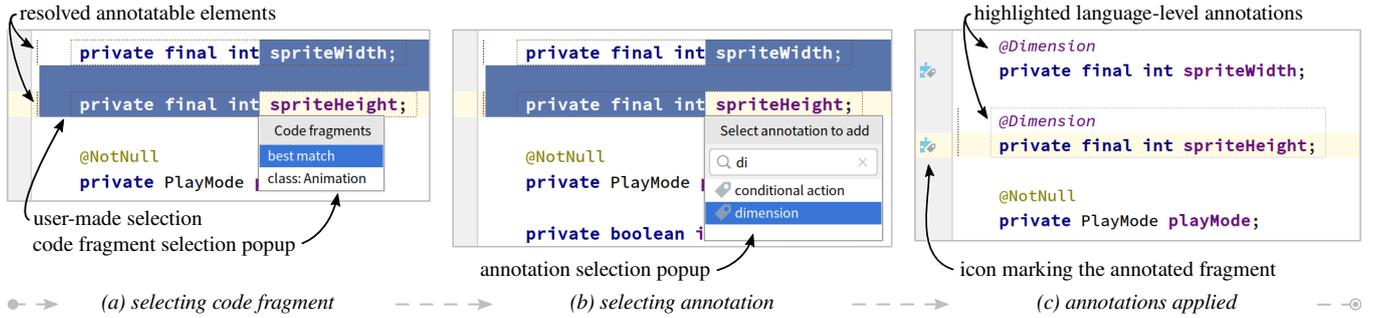


Fig. 3. Annotation process in *Connotator*. (a) The user selects code and uses action to resolve annotatable elements. (b) The user uses action to apply annotation to the selection and then selects desired annotation. (c) The tool applies *Java* LLAs because this type is configured for class fields in the project.

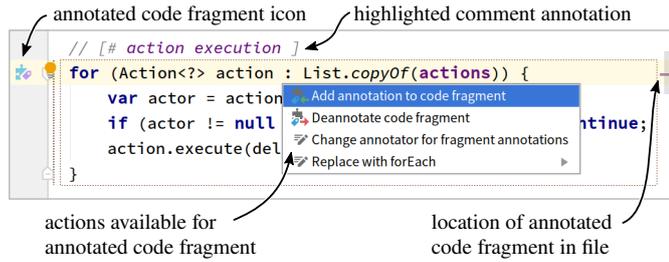


Fig. 4. IDE editor showing annotated block statement with associated actions and source code editor augmentations.

elements; in fact, compilers and interpreters for most languages ignore comments already in the lexing phase.

However, IDEs need to know about every concrete syntax tree token in order to be able to map each character from the file to the corresponding parsed node, and *vice versa*. For this reason they use custom lexers and parsers that preserve comments [12]. Nevertheless, only comment's parent can be determined from the tree (e.g., a comment inside a method), which is not enough to unambiguously assign comments to code elements. Does a comment standing alone on a first line within a method relate to the method (parent) or to a statement below? One possibility for dealing with such ambiguities is to deploy a set of conventions, or rules, to resolve them.

In the design of comment-to-element binding rules for a tool that needs to be able to find comments in the existing code, as well as to generate them when elements are annotated, we need to consider their following two properties:

- *Placement* of comments relative to elements they are bound to. There are many such relative placements that can be supported; see, e.g., the work of Sommerlad *et al.* [13] or our examples in Fig. 5.
- *Type* of comments that should be used. Particularly *end of line* and *block* comments are often supported by languages, sometimes complemented by conventional format of *documentary comment* (as *Java*'s *JavaDoc*).

Examples of possible relative comment-to-element placements and comment types are in Fig. 5. In the following, we describe how these placements and comment types are interpreted by *Connotator*, which supports their configuration

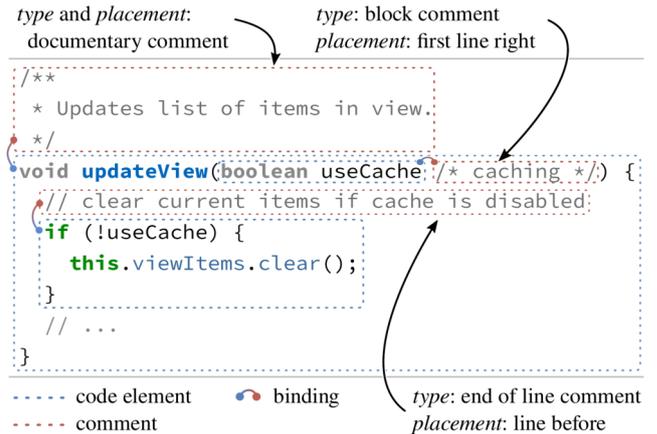


Fig. 5. Comment placements and types with bound code elements

for each annotatable element type.

- *End of line* comment is bound to the *if* statement on the next line through the *line before* relative placement.
- *Block comment* is bound to the method parameter *useCache* through the *first line right* relative placement. The *block* type of comment is required in this context because the bound element is followed by more tokens.
- *Documentary comment* is bound to the method *updateView* through the equally named *documentary comment* placement. This placement expresses the intention to bound the conventional documentary comment to the element it documents.

Because the user can manually insert a comment annotation at invalid position (such that does not bind it to any annotatable element), *Connotator* is able to issue warnings through in-editor highlighting when such comment is found.

### III. EXPERIMENT: ANNOTATING CODE WITH COMMENT ANNOTATIONS

Our main motivation behind the presented tool support for annotation process is in reducing overhead of annotating a source code. To evaluate any effects our prototype tool *Connotator* can have in this regard, we prepared a task in which a portion of a selected application's code base needs to be annotated with comments containing high-level concern

metadata. We specify what kinds of code elements can be annotated and where should the annotating comments be placed relatively to the annotatable elements: together, we will call this set of constraints the *annotation rules*.

In this section we present an experiment in which we compare two groups of students-programmers performing the described task. The first group uses *IntelliJ IDEA* with *Connotator* installed and the second, control group, uses *IntelliJ IDEA* in its default setup. We formulate the following two research questions for the experiment.

**RQ 1.** *Do programmers annotate source code with comments more consistently with defined annotation rules if they are guided by a metadata-aware tool that adheres to these rules?*

**RQ 2.** *Do programmers annotate source code with comments more quickly if the development environment is aware of the comment annotations?*

### A. Hypotheses

In the experiment we focus on two aspects of the annotation process: comment annotations *placement consistency* and *annotating speed*. In the following, we formulate related hypotheses. The goal of the experiment is to statistically test those hypotheses with a confidence level of 95% ( $\alpha = 5\%$ ).

1) *Comment Annotations Placement Consistency*: In *Connotator*, comment annotations placement rules are used to bind SCAs to specific code elements. We hypothesize that *Connotator* will help to *increase* comment annotations *placement consistency* in comparison to manual annotating. We define placement consistency (*PC*) of comment annotations as

$$PC = \frac{N_C}{N_A} \times 100\%$$

where  $N_C$  is the number of correctly (according to the annotation rules) placed comment annotations, and  $N_A$  is the number of all comment annotations placed during the annotating task. We formulate the following null and alternative hypotheses for RQ 1:

H1<sub>null</sub>: The **placement consistency** of comment annotations created during code annotation task with *Connotator* **is equal** (=) to the placement consistency of comment annotations created during the same annotation task with the standard *IntelliJ IDEA* setup.

H1<sub>alt</sub>: The **placement consistency** of comment annotations created during code annotation task with *Connotator* **is higher** (>) than the placement consistency of comment annotations created during the same annotation task with the standard *IntelliJ IDEA* setup.

2) *Annotating Speed*: *Connotator* abstracts the annotation process for different annotation types into a set of IDE actions. We hypothesize that these actions, when combined with the annotation rules for annotatable elements, will *increase* *annotating speed* of programmers performing code annotation task. We define annotating speed (*AS*) as

$$AS = \frac{N_A}{t}$$

where  $N_A$  is the number of all comment annotations placed in the code during the annotating task and  $t$  is the total time needed to complete the task. We formulate the following null and alternative hypotheses for RQ 2:

H2<sub>null</sub>: The **annotating speed** during code annotation task with *Connotator* **is equal** (=) to the annotating speed during the same annotation task with the standard *IntelliJ IDEA* setup.

H2<sub>alt</sub>: The **annotating speed** during code annotation task with *Connotator* **is higher** (>) than the annotating speed during the same annotation task with the standard *IntelliJ IDEA* setup.

### B. Setup

1) *Participants*: Participants of the experiment were 36 bachelor's degree Computer Science students from our department. They were in their fourth semester with programming courses and were familiar at least with languages *C* and *Java*. These students formed two study groups (not equally sized) of the Component Programming course, in which the *Java* language is used. One group of students was chosen as the experimental group where *Connotator* was used: we will call it the *Connotator* group. The other group of students was used as the control group working with standard *IntelliJ IDEA* IDE setup: the *manual* group. The *Connotator* group contained 20 participants and the *manual* group 16.

2) *Code for Annotation*: As a target for the annotating task we used source code of application for managing notes for bibliographic entries called *EasyNotes*.<sup>6</sup> It is a small-scale (about 2700 lines of code) project written in *Java*. An advantage of its code base is the presence of high-level concern annotations in a form of *Java* annotations, which were added by its author for the purpose of the study performed by Sulír *et al.* [4]. This provided us a very good starting point for preparing our own annotating task. From 25 available concern annotation types, we selected 10 that covered a large portion of the application's domain logic, its data model and persistence. We left out all the code directly related to the graphical user interface because it contained more complicated code generated by a UI designing application.

However, high-level concerns may be difficult to recognize in an unfamiliar code base. This difficulty of program comprehension represents the main confounding factor for our experiment because it can negatively affect the annotation speed (our dependent variable) and correctness of *contextual* placement of annotations (which we do not evaluate). Our attempt to minimize the influence of this factor was to try and bring the annotating close to a mechanical process, not unlike one performed by a programmer who is already familiar with the code. For this purpose, we renamed several identifiers to names that included some form of the relevant concern name.

3) *Format of Comment Annotations*: The two groups of participants did not use exactly the same comment structure

<sup>6</sup>Source code of the application *EasyNotes* is available at <https://github.com/MilanNosal/easy-notes>.

for annotations. As shown in Fig. 4, *Connotator* uses specific comment annotation format where the annotation name needs to be placed between prefix [ # and suffix ] within the text of a comment. These additional symbols are, however, meaningless without the tool and they would pose unnecessary hindrance for participants working without *Connotator*. For this reason, participants in the control group did use only simple prefix # before annotation name in comments to clearly designate that the following text is meant to be an annotation.

4) *Additional Materials*: We prepared two documents for study participants: annotation rules they need to follow during the task and a user guide for *Connotator*. We kept these documents short so they would fit each on one sheet of paper.

The document with annotation rules was designed to guide participants through the task of annotating EasyNotes' code. It described the form of comment annotations, their possible relative placements, and paired each type of annotatable element with required comment placement (for the last see Table I). The document was concluded with a table of actual annotations the participants should use. For each of the 10 annotations it specified words related to high-level concerns that could be found in element identifiers. It also explained the high-level meaning of each annotation. For example, for concern *citing* the table listed "*citing of publications*" as its explanation and "*cite, citation, publication*" as related words in identifiers.

In addition to the annotation rules, we provided participants in the *Connotator* group with a brief user guide of the tool. This guide explained the role of the annotations panel and the available workflows to select, annotate and deannotate code fragments. The guide also presented the *Connotator*'s ways of signalling through code highlighting whether a specific comment annotation is considered to be valid or invalid according to the configuration.

5) *Environment*: The experiment took place in our department's software laboratory room containing 20 computers with widescreen full HD displays and *IntelliJ IDEA 2017.3.5* installed on Windows 10 OS. We also set up a screen recording application to record the participants' annotating sessions for extracting the task completion times and for later analysis of their performance. We informed participants that their session was going to be recorded.

### C. Procedure

We carried out the experiment in two separate sessions, each for one group of participants and lasting 90 minutes (the

duration of a lab lesson). Each session proceeded as follows.

When the participants came for the experiment into the laboratory room, they already had their environment prepared: *IntelliJ IDEA* was running with the EasyNotes project opened and the screen recording was started. The *Connotator* group had the tool configured in accordance with the annotation rules.

First, the experimenter—the author of this paper—introduced the concept of annotating source code with high-level concerns. Then he presented the EasyNotes application, explaining its purpose and demoing its functionality.

In the next step, the experimenter handed over printouts of prepared materials, each labeled with a unique participant number. Then he walked the participants through its individual sections: form of comment annotations, possible comment placements, the comment annotation placements rules and the annotations themselves. For the *Connotator* group, the experimenter then covered the *Connotator* user guide complemented by presentation of its usage.

Next, the participants were asked to proceed with the task. They were also asked to minimize the IDE window and notify the experimenter when they finish. When each participant finished their task, he or she was asked to fill out a prepared questionnaire. At the end, the experimenter collected annotated projects and videos of annotating sessions.

### D. Evaluation

The first phase of evaluation consisted of analyses of captured videos, in which we needed to determine the actual duration of each participant's annotating task. Next, we proceeded with analysing annotated projects. The projects were analysed by counting created comment annotations. In every project annotated by an experiment participant we counted:

- Total number of annotations in comments.
- Number of invalid annotations, which were further categorized as comment annotation with:
  - *invalid placement*: annotations occurrences in comments that, given their placement in the code, did not annotate any element according to the annotation rules,
  - *invalid syntax*: comment annotations that did not use required syntax,
  - *invalid context*: comment annotations that annotated elements not related to the concerns expressed by these annotations.

### E. Results

Our data samples were unpaired, as each participant was either in *Connotator* or in *manual* group. There was one independent variable—availability of the *Connotator* tool—with nominal values "available" and "unavailable". Our dependent variables were comment annotations placement consistency and annotating speed, none of which looked normally distributed. We used the *Mann-Whitney U test* as a statistical test for our hypotheses. We report the results in the following.

TABLE I  
ANNOTATION RULES FOR THE ANNOTATING TASK

Annotatable code fragment type	Comment annotation placement
Class	
Method	Documentation comment
Class field	
Simple statement	
Method parameter	First line right
Block statement	Line before

1) *Comment Annotations Placement Consistency*: The placement consistency was higher in the *Connotator* group. The mean value in this group was 99.59%, compared with the mean of 83.55% in the *manual* group. Only 3 comment annotations were placed incorrectly for the *Connotator* group (participants typed them manually and ignored the tool’s warnings). On the other hand, *manual* group had 2 extreme outliers, who reached PC of only 46.43% and even 0.0%, respectively. If we exclude these two participants, the mean for the *manual* group rises to 92.17%. Statistical results are summarized as box plots in Fig. 6.

With or without the two extreme cases in the *manual* group, the computed  $p$  value is well below 0.001 ( $6.42 \times 10^{-7}$  and  $1.74 \times 10^{-6}$  respectively), which is also below our significance level (0.05). Thus, we reject  $H_{1null}$  and accept  $H_{1alt}$ . The conclusion is that comment annotations placement consistency was higher in the *Connotator* group and the result is statistically significant.

2) *Annotating Speed*: The annotating speed was also higher in the *Connotator* group, with the mean of 2.07 annotations per minute. The *manual* group reached the mean of 1.48 annotations per minute. See Fig. 7(a) for the box plot.

The computed  $p$  value for annotating speed is 0.11, which is above our significance level (0.05). Thus, we fail to reject  $H_{2null}$ . The conclusion is that while annotating speed was higher in the *Connotator* group, the result is **not** statistically significant.

It is, however, interesting to note that when we consider just the duration of the annotating sessions (box plot shown in Fig. 7(b)), the difference between groups is more prominent: participants from the *Connotator* group finished their task in 42.64 minutes in average, while for *manual* group the average is 52.76 minutes. The computed  $p$  value for session durations is  $< 0.001$ . We discuss the possible reasons for this

discrepancy between annotating speed and annotating session duration (among other observations) in Section IV.

### F. Threats to Validity

In the following, we discuss threats to the validity of the experiment and relevant control actions taken.

a) *Internal validity*: Assignment of participants into groups was not strictly random: we used existing groups of students, in which randomness is not guaranteed. The alternative was to randomly assign half of each study group to the experimental group and the other half to the control group. However, in such arrangement, the experimenter would need to present the annotating tool in front of participants assigned to the control group (they would be in the same room), which could also have an effect on the result.

Pilot-testing was limited to one participant who performed the annotating task with *Connotator*. At that time, 13 high-level annotations were selected. As we considered the time needed to reasonably complete the task too long, we decided to lower the count of annotations to 10.

b) *External validity*: Participants of our experiment were students, not professional programmers. According to the findings of Salman *et al.* [14], it might not have great effects on the results, because the tested approach—the *Connotator* tool—is new for both students and professionals. Nevertheless, we would need code authors or maintainers (who would know the project in detail) to eliminate the effect of program comprehension on the result. We attempted to eliminate this effect by making the task more mechanical, as described in Section III-B2.

Within the broader approach of using extrinsic metadata for program comprehension, we tested only its part—the annotation process. Without further integration of bound metadata into the IDE, their presence in annotations is not well utilized. However, the annotation process is the most time-consuming part of the approach, and we strive for a better supporting tool.

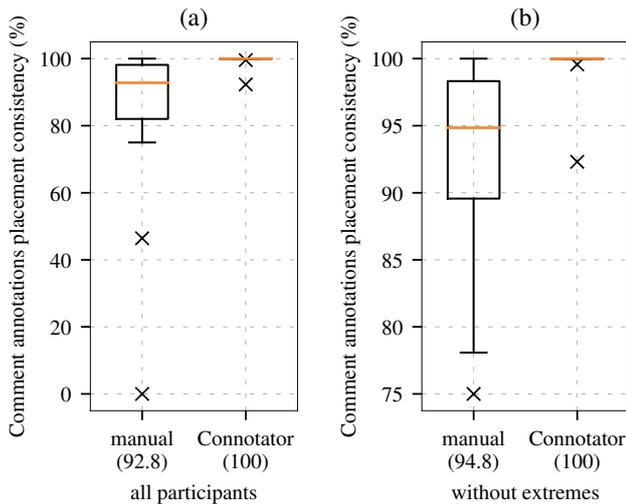


Fig. 6. The results of statistical evaluation of differences between groups of participants regarding their comment annotations placement consistency: (a) with all participants, (b) with the two outliers (0% and 46.4% PC) from the *manual* group removed. Median values are included below group names.

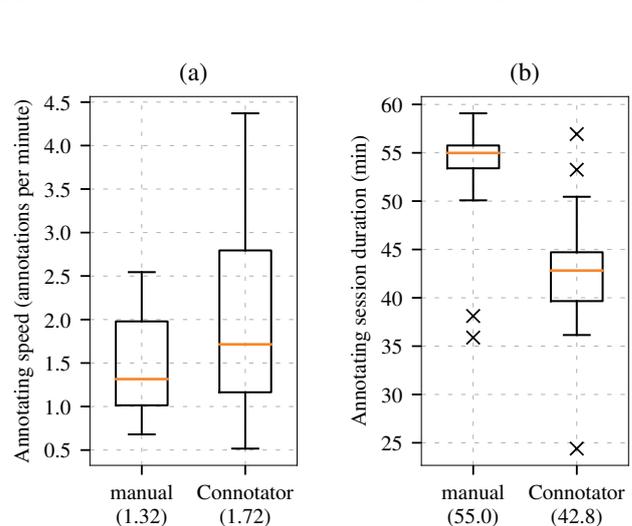


Fig. 7. The results of statistical evaluation of differences between groups of participants regarding (a) their annotating speed and (b) duration of their annotation sessions. Median values are included below group names.

c) *Conclusion validity.*: The most prominent threats to conclusion validity are the small number of subject participating in the experiment (36) and the confounding factor of needed comprehension of annotated source code.

#### IV. DISCUSSION OF THE EXPERIMENT

In this section, we present observations regarding the data we obtained by analysing source codes annotated in the experiment.

##### A. Differences in Number of Created Annotations

Based on our version of annotated EasyNotes' source code, which itself was based on annotations from its author and refined to finer granularities allowed by SCAs, we consider 70 to 90 comment annotations for an optimal result of annotating. Participants annotated code on the basis of the annotations table that we provided them. Ultimately, the specific code elements—and their count thereof—that participants chose to annotate depended on their understanding of both the annotations and of the code. In processed projects, we saw that the numbers of annotations created during the task varied significantly.

From the box plot in Fig. 8, we can see that the *manual* group performed better with regard to the number of annotations. The median (72.5 annotations) is closer to our optimal count than the median of *Connotator* group (61.5 annotations) and there are no very low (<30) nor very high (>200) values.

The most frequently and most inconsistently used was annotation *domain entity*. Some participants took this annotation too broadly and annotated a majority of variables named `note` or `notes`. Interestingly, the extreme cases of such very general understanding of this annotation (almost 100 occurrences in the project) were present only in the *Connotator* group.

Described differences in number of created annotations, especially the tendency towards lower count of annotations in the *Connotator* group, may be behind the discrepancy between annotating speed and session duration distributions (see Fig. 7). Also, 3 participants in the *Connotator* group annotated code in less than 40% of files they were asked to annotate, in comparison to only one such participant in the *manual* group.

We conclude that in order to more reliably assess the effect of our tool on annotating speed, the task should more precisely define both the elements to be annotated and the determining factor of when the task can be considered as completed. This may reduce the variability in annotations counts or at least allow us to exclude clearly incomplete tasks.

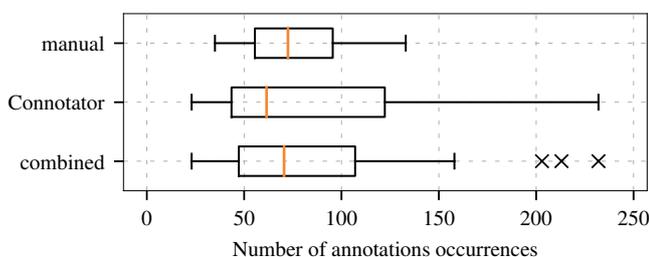


Fig. 8. Distributions of number of created comment annotations. Shows distributions for both groups and for all participants combined.

##### B. A Closer Look at Placement Consistency

Comment annotations placement consistency (*PC*) showed to be significantly higher in the *Connotator* group (see Section III-E1). This result may be not surprising as this group used tool that prevented *misplaced*<sup>7</sup> comment annotations. Our interest was, however, to find out how many misplaced annotations there would be in the *manual* group and whether the difference would be significant.

Fig. 9 shows absolute numbers of misplaced annotations in relation to all annotations created by individual participants. Only one participant in the *manual* group managed to make no placement mistakes, but he made totally only 35 annotations (the lowest number in the group).

##### C. Invalid Syntax or Context of Created Comment Annotations

There were 0 comments with invalid syntax in the *Connotator* group. *Manual* group made syntactic mistakes in 1.6% of comments in average, mainly by omitting the # prefix. Three participants in the *Connotator* group misspelled name of one annotation, which resulted in having them misspelled at every occurrence in the code. However, due to *Connotator*'s annotation renaming feature, such issue can easily be fixed. On the other hand, *manual* group had 2.6% of comments in average with misspelled annotation names. Without a supporting tool, such errors lead to inconsistencies that hinder the usage of such *comment tags* with common search tools [15].

As the tested annotation process does not influence program comprehension, we expected to find no difference in the numbers of contextually invalid annotations. This expectation was confirmed as these was only small and statistically insignificant difference: median values for percentages of contextually invalid annotations were 13.6% and 11.7% for *Connotator* and *manual* groups, respectively.

##### D. Observations from the Questionnaire

In the questionnaire we asked participants questions about how they would rate their understanding of annotations

<sup>7</sup>Misplaced comment annotation is an annotation within a comment that is not bound to any code element because its placement or comment type does not match annotation rules.

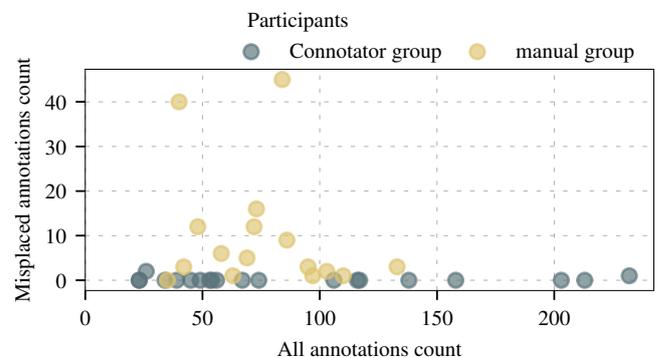


Fig. 9. Number of misplaced annotations of each participant.

meaning and of EasyNotes source code on a 5-point scale. Generally, participants expressed that they understood meanings of annotations and that they were able to comfortably navigate source code of EasyNotes. Differences between groups were minor, with marginally higher (i.e., better understanding) means in the *manual* group.

Next, we asked question regarding the annotation process, in which the participants answered as follows.

- 45% of participants from the *Connotator* group stated that annotating was simple and fast, compared with only 12.5% in the *manual* group.
- 43.8% of participants from the *manual* group considered annotating as laborious and 62.5% stated they copied existing comment annotations to create new ones.
- 90% of participants from the *Connotator* group considered the annotating tool as helpful.

We also asked how often they needed to check the document with annotation rules (on a 10-point scale from “in 1 out of 10 cases” to “in 10 out of 10 cases”). The responses are plotted in Fig. 10 and show that participants in the *Connotator* group reported less frequent usage of annotation rules (median of 3/10) than in the *manual* group (with median of 5.5/10). These responses indicate that participants working with our tool were less occupied by the details of the annotation process.

## V. RELATED WORK

In this section we look at other approaches and tools that share similarities with ours presented in this paper.

Mattis *et al.* present an approach named *Concept-Aware Programming Environment* [16]. They are interested in making programming environments aware of *concepts* that are present in the code through identifier names, with the goal to help programmers to build their mental model of the code. Another use-case is in detecting *architectural drift*: change in meanings and distributions of words used in names of identifiers during program evolution. Their method for finding concepts in the code is automated, but allows programmer’s intervention and correction. For sharing of corrected concepts in a distributed workspace, they suggest to embed them in comments, which would result in SCAs conceptually similar to our ones. Integration of their approach into development environments consists of tools for concept exploration, custom class diagrams, and concept-augmented IDE editor, debugger and VCS tools.

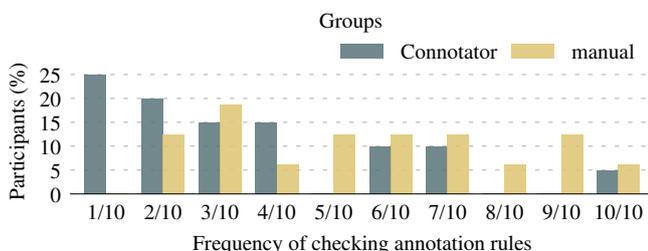


Fig. 10. The frequency of checking annotation rules by participants in  $n$  out of 10 cases of inserting an annotation into the code.

We presented a uniform annotation process that works for all annotation types and, considering their limitations, allows to choose the appropriate type per use case. Cazzola *et al.* extended LLAs for languages *C#* [17] and *Java* [18] to finer granularities through customized compilers, by which they extended the applicability of this single annotation type to code blocks and expressions. In comparison with our approach, the availability of annotations in the compiler and for processing through reflection at runtime is an advantage. However, a special tool (in this case a custom compiler) is still needed, and each such solution is restricted to a single language.

Current implementation of *Connotator* has only a simple metadata model, where annotations can represent high-level concerns through their names. A practical extension of this model would be to link external resources to the metadata. Similar thing was done by Baltes *et al.* who designed *SketchLink* tool [5] for linking sketches documenting high-level design to code elements through SCAs. Their design included a service for uploading and managing images of sketches from mobile devices and web browsers. IDE plug-in linked these images through unique identifier included in source code comments.

Our approach to select structurally valid annotatable code fragments uses AST patterns. Kästner *et al.* [19] and Behringer *et al.* [20] use AST rules to ensure structurally valid separation of features in feature-oriented development of software product lines. Kästner *et al.* call the annotation process *coloring* and Behringer *et al.* extend it with *snippet* code organization system for managing feature variability.

Cséri *et al.* [21] present their approach to assign source code comments to specific elements of the AST. They were interested in comment-to-element assignment for software maintenance tool, but had to work with legacy codebases, inside of which the assignment needed to happen. Their solution, similar to ours, consists of project-specific rules for defining relative comment placements, but the rules are more complex, supporting, e.g., assignment of a single comment node to multiple code elements. On the other hand, they do not differentiate types of comments, because they only process existing comments and do not need to generate new ones.

Rule-based comment assignment is also used in tool *TagSEA* by Storey *et al.* [22]. It uses a simple rule: comments are bound to the closest enclosing *Java* element. Such rule is sufficient if metadata granularity does not go below methods.

In contrast to our per-project configurable comment-to-element assignment, Sommerlad *et al.* [13] used fully-automatic assignment, distinguishing *leading*, *trailing* and *freestanding* comments. Their goal was to retain all comments and their positions while refactoring the code.

We used annotations for high-level metadata from the problem domain. Sulír and Porubán in their approach [23] used annotations to preserve low-level runtime information.

## VI. CONCLUSION AND FUTURE WORK

In this paper, we presented our work towards allowing programmers to more easily preserve their high-level knowledge of source code they create by annotating it.

We gave an overview of the concept of the metadata-aware development environment and focused on its first building block: the support for the annotation process. For annotating code with three different types of annotations, and making the annotating workflow uniform, we designed an abstraction of the process. First, a programmer chooses source code elements to annotate and annotation representing metadata that should be bind to these elements. Then, a specific annotation is applied automatically by a supporting tool. We also described our prototype of such a tool in the form of a plugin for *IntelliJ* platform-based IDEs, called *Connotator*.

Finally, we reported the experiment in which we evaluated *Connotator* regarding its effect on the annotation process. We confirmed the hypothesis that the tool could increase placement consistency of comment annotations with annotation rules. The group of participants using the tool achieved higher consistency and reported less distraction by the details of the annotating than the group without the tool. The hypothesis that the tool increases annotating speed was not confirmed, although the group using the tool tended to finish the task sooner.

The natural next progress is to explore in detail the remaining two aspects of MADE to better utilize the preserved metadata and facilitate program comprehension. An interesting direction may also be in merging intrinsic metadata already available in IDEs with the preserved, extrinsic ones, and providing a querying facility for the resulting model. The queries could be used to customize views of code provided by an IDE, similarly to the concept of *scriptable* IDE presented by Asenov *et al.* [24].

Although our approach supports multiple types of annotations, we used only one type in the presented experiment. We will focus on assessing the annotation process using a mixture of annotation types in future evaluations.

## REFERENCES

- [1] T. D. LaToza and B. A. Myers, "Hard-to-answer questions about code," in *Evaluation and Usability of Programming Languages and Tools on - PLATEAU '10*. ACM Press, oct 2010. doi: 10.1145/1937117.1937125 pp. 1–6.
- [2] V. Vranić, J. Porubán, M. Bystrický, T. Frt'ala, I. Poláček, M. Nosál', and J. Lang, "Challenges in Preserving Intent Comprehensibility in Software," *Acta Polytechnica Hungarica*, vol. 12, no. 7, pp. 57–75, 2015. doi: 10.12700/APH.12.7.2015.7.4
- [3] B. Dit, M. Revelle, M. Gethers, and D. Poshyvanyk, "Feature location in source code: a taxonomy and survey," *Journal of Software: Evolution and Process*, vol. 25, no. 1, pp. 53–95, jan 2013. doi: 10.1002/smr.567
- [4] M. Sulír, M. Nosál', and J. Porubán, "Recording concerns in source code using annotations," *Computer Languages, Systems and Structures*, vol. 46, pp. 44–65, nov 2016. doi: 10.1016/j.cl.2016.07.003
- [5] S. Baltes, P. Schmitz, and S. Diehl, "Linking sketches and diagrams to source code artifacts," in *Proceedings of the 22nd ACM SIGSOFT International Symposium on Foundations of Software Engineering - FSE 2014*. New York, New York, USA: ACM Press, 2014. doi: 10.1145/2635868.2661672 pp. 743–746.
- [6] M. Nosál', J. Porubán, and M. Nosál', "Concern-oriented source code projections," in *Proceedings of the 2013 Federated Conference on Computer Science and Information Systems*, Kraków, 2013, pp. 1541–1544.
- [7] W. Ji, T. Berger, M. Antkiewicz, and K. Czarniecki, "Maintaining feature traceability with embedded annotations," in *Proceedings of the 19th International Conference on Software Product Line - SPLC '15*. New York, New York, USA: ACM Press, 2015. doi: 10.1145/2791060.2791107 pp. 61–70.
- [8] J. Juhár and L. Vokorokos, "Exploring code projections as a tool for concern management," *Acta Electrotechnica et Informatica*, vol. 16, no. 3, pp. 26–31, 2016. doi: 10.15546/aei-2016-0020
- [9] K. Rástočný and M. Bieliková, "Metadata Anchoring for Source Code: Robust Location Descriptor Definition, Building and Interpreting," in *24th International Conference on Database and Expert Systems Applications*. Prague: Springer, Berlin, Heidelberg, 2013. doi: 10.1007/978-3-642-40173-2\_30 pp. 372–379.
- [10] M. P. Robillard and F. Weigand-Warr, "ConcernMapper: simple view-based separation of scattered concerns," in *Proceedings of the 2005 OOPSLA workshop on Eclipse technology eXchange - eclipse '05*. New York, New York, USA: ACM Press, oct 2005. doi: 10.1145/1117696.1117710 pp. 65–69.
- [11] M. Sulír, M. Bačíková, S. Chodarev, and J. Porubán, "Visual augmentation of source code editors: A systematic mapping study," *Journal of Visual Languages & Computing*, vol. 49, pp. 46–59, dec 2018. doi: 10.1016/J.JVLC.2018.10.001
- [12] D. Jemerov, "Implementing refactorings in IntelliJ IDEA," in *Proceedings of the 2nd Workshop on Refactoring Tools - WRT '08*. ACM Press, oct 2008. doi: 10.1145/1636642.1636655 pp. 1–2.
- [13] P. Sommerlad, G. Zraggen, T. Corbat, and L. Felber, "Retaining comments when refactoring code," in *Companion to the 23rd ACM SIGPLAN conference on Object oriented programming systems languages and applications - OOPSLA Companion '08*. New York, New York, USA: ACM Press, 2008. doi: 10.1145/1449814.1449817 p. 653.
- [14] I. Salman, A. T. Misirli, and N. Juristo, "Are Students Representatives of Professionals in Software Engineering Experiments?" in *2015 IEEE/ACM 37th IEEE International Conference on Software Engineering*. IEEE, may 2015. doi: 10.1109/ICSE.2015.82 pp. 666–676.
- [15] A. T. T. Ying, J. L. Wright, S. Abrams, A. T. T. Ying, J. L. Wright, and S. Abrams, "An exploration of how comments are used for marking related code fragments," in *Proceedings of the 2005 workshop on Modeling and analysis of concerns in software - MACS '05*, vol. 30, no. 4. New York, New York, USA: ACM Press, 2005. doi: 10.1145/1083125.1083141 pp. 1–4.
- [16] T. Mattis, P. Rein, S. Ramson, J. Lincke, and R. Hirschfeld, "Towards concept-aware programming environments for guiding software modularity," *Proceedings of the 3rd ACM SIGPLAN International Workshop on Programming Experience*, pp. 36–45, 2017. doi: 10.1145/3167110
- [17] W. Cazzola, A. Cisternino, and D. Colombo, "Freely annotating C#," *Journal of Object Technology*, vol. 4, no. 10, pp. 31–48, 2005. doi: 10.5381/jot.2005.4.10.a2
- [18] W. Cazzola and E. Vacchi, "@Java: Bringing a richer annotation model to Java," *Computer Languages, Systems and Structures*, vol. 40, no. 1, pp. 2–18, 2014. doi: 10.1016/j.cl.2014.02.002
- [19] C. Kästner, S. Apel, and M. Kuhlemann, "Granularity in software product lines," in *Proceedings of the 13th international conference on Software engineering - ICSE '08*. New York, New York, USA: ACM Press, 2008. doi: 10.1145/1368088.1368131 p. 311.
- [20] B. Behringer, L. Kirsch, and S. Rothkugel, "Separating features using colored snippet graphs," in *Proceedings of the 6th International Workshop on Feature-Oriented Software Development - FOSD '14*. New York: ACM Press, 2014. doi: 10.1145/2660190.2660192 pp. 9–16.
- [21] T. Cséri, Z. Szügyi, and Z. Porkoláb, "Rule-based assignment of comments to AST nodes in C++ programs," in *Proceedings of the Fifth Balkan Conference in Informatics - BCI '12*. New York, New York, USA: ACM Press, 2012. doi: 10.1145/2371316.2371381 pp. 291–294.
- [22] M.-A. Storey, L.-T. Cheng, I. Bull, and P. Rigby, "Shared waypoints and social tagging to support collaboration in software development," in *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work - CSCW '06*. New York, New York, USA: ACM Press, 2006. doi: 10.1145/1180875.1180906 pp. 195–198.
- [23] M. Sulír and J. Porubán, "Exposing Runtime Information through Source Code Annotations," *Acta Electrotechnica et Informatica*, vol. 17, no. 1, pp. 3–9, 2017. doi: 10.15546/aei-2017-0001
- [24] D. Asenov, P. Müller, and L. Vogel, "The IDE as a scriptable information system," in *Proceedings of the 31st IEEE/ACM International Conference on Automated Software Engineering - ASE 2016*. New York, New York, USA: ACM Press, 2016. doi: 10.1145/2970276.2970329 pp. 444–449.

# 10<sup>th</sup> Workshop on Scalable Computing

**T**HE world of large-scale computing continuously evolves. The most recent addition to the mix comes from numerous data streams that materialize from exploding number of cheap sensors installed “everywhere”, on the one hand, and ability to capture and study events with systematically increasing granularity, on the other. To address the needs for scaling computational and storage infrastructures, concepts like: edge, fog and dew computing emerged.

Novel issues involved in “pushing computing away from the center” did not replace open questions that existed in the context of grid and cloud computing. Rather, they added new dimensions of complexity and resulted in the need of addressing scalability across more and more complex ecosystems consisting of individual sensors and micro-computers (e.g. Raspberry PI based systems) as well as supercomputers available within the Cloud (e.g. Cray computers facilitated within the MS Azure Cloud).

Addressing research questions that arise in individual “parts” as well as across the ecosystem viewed from a holistic perspective, with scalability as the main focus is the goal of the Workshop on Scalable Computing. In this context, the following topics are of special interest (however, this list is not exhaustive).

## TOPICS

- General issues in scalable computing
  - Algorithms and programming models for large-scale applications, simulations and systems
  - Large-scale symbolic, numeric, data-intensive, graph-oriented, distributed computations
  - Fault-tolerant and consensus techniques for large-scale computing
  - Resilient large-scale computing
  - Data models for large-scale applications, simulations and systems
  - Large-scale distributed databases
  - Load-balancing / intelligent resource management in large-scale applications, simulations and systems
  - Performance analysis, evaluation, optimization and prediction
  - Scientific workflow scheduling
  - Data visualization
  - On-demand computing
  - Virtualization supporting computations
  - Volunteer computing
  - Scaling applications from small-scale to exa-scale (and back)
  - Big data real-time computing / analytics

- Economic, business and ROI models for large-scale applications
- Emerging technologies for scalable computing
  - Cloud / Fog / Dew computing architectures, models, algorithms and applications
  - High performance computing in Cloud / Fog / Dew
  - Green computing in Cloud / Fog / Dew
  - Performance, capacity management and monitoring of Cloud / Fog / Dew configuration
  - Cloud / Fog / Dew application scalability and availability
  - Big Data cloud services
  - Architectures for large-scale computations (GPUs, accelerators, quantum systems, federated systems, etc.)
  - Self\* and autonomous computational / storage systems

## EVENT CHAIRS

- **Ganzha, Maria**, Warsaw University of Technology, Poland
- **Gusev, Marjan**, University Sts Cyril and Methodius, Macedonia
- **Paprzycki, Marcin**, Systems Research Institute Polish Academy of Sciences, Poland
- **Petcu, Dana**, West University of Timisoara, Romania
- **Ristov, Sashko**, University of Innsbruck, Austria

## PROGRAM COMMITTEE

- **Barbosa, Jorge**, University of Porto, Portugal
- **Camacho, David**, Universidad Autonoma de Madrid, Spain
- **Carretero, Jesus**
- **D’Ambra, Pasqua**, IAC-CNR, Italy
- **Durillo, Juan**, Leibniz Supercomputer of the Bavarian Academy of Sciences and Humanities, Germany
- **Garcia Valdez, Mario**
- **Gordon, Minor**, Software development consultant, United States
- **Gravvanis, George**, Democritus University of Thrace, Greece
- **Grosu, Daniel**, Wayne State University, United States
- **Holmes, Violeta**, The University of Huddersfield, United Kingdom
- **Kalinov, Alexey**, Cadence Design Systems, Russia
- **Kecskemeti, Gabor**, Liverpool John Moores University, United Kingdom
- **Kitowski, Jacek**, AGH University of Science and Technology, Department of Computer Science, Poland
- **Knepper, Richard**, Indiana University, United States

- **Lang, Tran Van**, Vietnam Academy of Science and Technology, Vietnam
- **Lastovetsky, Alexey**, University College Dublin, Ireland
- **Margaritis, Konstantinos G.**, University of Macedonia, Greece
- **Nosovic, Novica**, Faculty of Electrical Engineering, University of Sarajevo, Bosnia and Herzegovina
- **Pawłowski, Wiesław**, University of Gdańsk, Poland
- **Prodan, Radu**, University of Klagenfurt, Austria
- **Schikuta, Erich**, University of Vienna, Austria
- **Schreiner, Wolfgang**, Johannes Kepler University Linz, Austria
- **Shen, Hong**, University of Adelaide, Australia
- **Telegin, Pavel**, JSCC RAS, Russia
- **Tudruj, Marek**, Inst. of Comp. Science Polish Academy of Sciences/Polish-Japanese Institute of Information Technology, Poland
- **Vazhenin, Alexander**, University of Aizu, Japan
- **Wei, Wei**, School of Computer science and engineering, Xi'an University of Technology, China
- **Wyrzykowski, Roman**, Czestochowa University of Technology, Poland
- **Zavoral, Filip**, Charles University, Czech Republic

# Measure of Adequacy for the Supercomputer Job Management System Model

Anton Baranov, Dmitriy Lyakhovets, Gennady Savin, Boris Shabanov, Pavel Telegin  
Joint Supercomputer Center of the Russian Academy of Sciences  
Leninskiy pr., 32a, Moscow, Russia  
Email: {abaranov, ptelegin, shabanov}@jscc.ru, anetto@inbox.ru

**Abstract**—In this paper we investigate the problem of modelling modern supercomputer job management systems (JMS). When modelling the JMS, one of the main issues is the adequacy of the model used in experimental studies. The paper attempts to determine the measure of the JMS model adequacy by comparing the characteristics of two job streams, one of which was acquired from a real supercomputer and the other is obtained from the JMS model. We show that the normalized Euclidean distance between vectors of jobs residence times obtained from the job streams of the real system and the JMS model can serve as a measure of the adequacy of the JMS model. The paper also defines the reference value of the measure of adequacy corresponding to the JMS model with virtual nodes.

## I. INTRODUCTION

SUPERCOMPUTER centers are usually shared facilities for the users. The users share the supercomputer computational field, which consists of computational nodes (CN) integrated by a high-performance communication network. Typically, to perform calculations on a supercomputer, the user must create a so-called passport of computational job. The passport consists of a parallel program, input data and system requirements (number of cores or nodes, amount of RAM) and execution time limit.

Special software [1] like SLURM [2], PBS [3] or the Russian native job management system SUPPZ [4] manage jobs in supercomputers. The kernel of any job management system (JMS) is the scheduler. The scheduler generates a schedule for the jobs launches according to job passports. Information in the job passport includes required execution time of job, amount and types of resources. The JMS scheduler provides quite an accurate time prediction of the launch time for each queued job. Changes to this forecast are usually made when the schedule is renewed due to a new job submission, job removal from the queue, or premature completion of a running job.

A set of indicators is used to measure the quality of scheduler. These indicators include average load, average waiting time for a job in a queue, etc. [5]. These indicators are influenced by both the configuration parameters of the scheduler and the characteristics of the input job stream. At the same time, this influence is not always evident and cannot be esti-

mated or predicted, since modern JMS are rather complicated systems with many adjustable parameters. This is why JMS modelling is relevant for studying the way that the input stream characteristics and JMS configuration influence job scheduling quality indicators.

Functioning of the real system and its simulation will be somewhat different, this results two interrelated problems. First, it is necessary to find out how to measure the model reproduction accuracy of the simulated system, i.e. to determine the measure of accuracy (adequacy) of the model. This will make it possible to compare different models by their adequacy. Secondly, it is necessary to establish the maximum valid value (limit) of the measure of accuracy. The overrun of this limit means that the model is not adequate and cannot be used to analyse the real system behaviour. The main goal of the paper is search and selection of models adequacy measure, as well as definition of the adequacy limits for the JMS models.

## II. THE PROBLEM OF A SUPERCOMPUTER JOB MANAGEMENT SYSTEM MODELLING

A number of external and internal events occur during the operation of JMS. External events include job submission, premature job termination, deletion of job from a queue or job interruption by a user or an administrator, JMS start and stop, change in number of available computing nodes. The internal events include the next job launch at the appointed time according to the schedule. JMS logs the time stamp and type of each event.

We consider JMS simulation as the process of submitting the external events stream to the model input and logging the internal events stream. The model's resulting stream of the internal events should be similar to the same stream in the real system. These two streams are identical when the model is fully adequate.

Existing methods of JMS modelling can be categorized as follows:

1. Development of a JMS analytical model.
2. Experiment with a real supercomputer.
3. Study of the JMS with virtual nodes (VN) [6].
4. Development of a JMS simulation model.

This work was supported by RFBR grants no 18-29-03236 and 18-07-01325 and state assignment topic No. 0065-2019-0016

The analytical model allows investigating the impact of JMS changes on its interval indicators, but does not provide a way for predicting the launch time of individual jobs, which is necessary for forecasting. Due to the complexity of construction and orientation on interval indicators, the analytical model is not be considered in this paper.

### III. NATURAL EXPERIMENT AS THE WAY TO SIMULATE A JMS

By term “natural experiment”, we mean the reproduction of an input external event stream in a real supercomputer. Therefore, the JMS model in a natural experiment will be fully adequate. Nevertheless, a natural experiment cannot provide reproduction of simulation results with 100% accuracy. In fact, processing time of a job consists of three generally random variables:

- job launch time: the time spent by the JMS for the allocation of computational nodes and their configuration in accordance with the job requirements;
- job execution time on the selected nodes;
- job completion time: the time spent by the JMS to release the selected nodes, including control the completion of all job processes, deletion of temporary files and shared resources created by the job, reconfiguration of the nodes, etc.

Job launch and completion time will be referred to as overheads. The billing subsystems for the most of the JMS include overheads into job execution time. At the same time, the proportion of overheads is a random value and can depend on many factors, such as network delays, changes in the state of calculations in the operating system kernel, etc.

The main disadvantage of a natural experiment is difficulty of its reproduction, since expensive supercomputer resources in such an experiment will duplicate the calculations already performed. Practically, a natural experiment is performed by changing JMS studied parameters. In accordance with the change of the JMS quality indicators, the decision is made whether to save the changes or to return to the previous version of the JMS settings.

### IV. SIMULATION OF JMS WITH VIRTUAL NODES

Virtual nodes (VN) can be used to model the JMS. This is a software subsystem, which, instead of launching jobs on computational nodes of a computational field, makes a note that virtual nodes are engaged for the duration of the assignment. Real calculations are not performed in this case.

There are two ways to simulate a JMS with VN: in real time mode and in model time mode. In real-time VN is presented to the JMS as a computational field, the JMS actually operates in a natural experiment mode without launching jobs on a supercomputer. This allows us to speak about the accuracy of such modelling as comparable with the accuracy of a natural experiment. The disadvantage of this method is a long simulation time corresponding to the real time of the JMS operation.

The basis of JMS with virtual nodes in the model time is the idea of «advancing» system time in those moments when external or internal events do not occur. For example, if at some point of the experiment no new jobs are received, at the current moment, one job is being processed and it will be completed in an hour, then it is possible to move the system time one hour forward. Simulation in this case is significantly accelerated. To implement this method, it is necessary to develop a special software tool for advancing the system time with additional verification of the experimental results accuracy.

### V. JMS SIMULATION MODELLING

To build a JMS simulation model, specialized languages can be used, like AnyLogic, ExtendSIM, Simulink [7], GPSS World [8]. Modelling languages fully provide the modelling process – the model time advancing and the interaction of objects in the system, allowing the researcher to focus on the description of the essential properties and characteristics of the simulation model.

Beside specialized modelling languages, there are so-called JMS simulators: GridSim [9], CloudSim [10], WorkflowSim [11]. Simulators supply with a set of implemented job scheduling algorithms and provide the formation of interval indicators based on the processing of the input event stream. It is also necessary to mention JMS emulators, e.g. MicroGrid [12]. A distinctive feature of the emulator is the possibility of sharing the real system components and the emulated JMS parts in the experiment.

Existing simulation tools allow us to build a predictive JMS model and conduct experiments with it on any model input event stream. However, it is necessary to validate the experiments results for simulation models in order to determine the model adequacy. To do this, it is necessary to set a measure of adequacy, express this measure by some quantitative characteristic and determine the allowable limits of this characteristic values, within which the model will be considered adequate.

### VI. JMS EVENT STREAM MODEL

Let all events in the JMS occur at discrete points in time  $t_i$ . Consider the stream of independent submitted jobs  $J_1, J_2, \dots, J_k, \dots, J_N$ . Each job  $J_i$  in the queue has the following characteristics:

- the moment of the job submit  $r_i$ ;
- the required resources  $p_i$ ;
- ordered processing time  $e_i$ ;
- real processing time  $w_i$ ,  $0 \leq w_i \leq e_i$ , which consists of job launch time  $a_i$ , execution time  $b_i$ , completion time  $c_i$ .

Note that the actual execution time is not available for the job management system and cannot be used to build a schedule. As shown above,  $a_i$ ,  $b_i$  and  $c_i$  are random variables and can vary from launch to launch of the same job.

The scheduler determines the job launch moment  $s_i$ . Derived characteristics of the job are wait time for a job in the queue  $q_i = s_i - r_i$ ; job residence time (full time spent in the system from submit to job completion)  $f_i = q_i + w_i$ ; the moment of the job completion  $g_i = s_i + w_i$ .

An events stream with some characteristics is fed to the JMS model input. The result of the JMS model is an output model stream of events with a different characteristic set. Denote the characteristics of this stream in capital letters.

There are three well-established approaches to the formation of the input event stream [13]. The first approach is to use the real JMS event log. The approach allows reproducing the input event stream of a real supercomputer, taking into account all its features. The second approach is based on the SWF (Standard Workload Format) [14]. Event logs of some supercomputers, including university ones, published in SWF. The essential drawback is the incompleteness of the event flow: SWF represents only events related with jobs in the queue, and there is no information about changes in the nodes number, job deletions from the queue or interruptions in the job execution by the user. The third approach is to generate an input stream of events [15]. Each job parameter (submit time, ordered and real execution time, required computing resources) is a random variable with a certain distribution law. The law and distribution parameters are selected, as a rule, based on the analysis of the studied supercomputers event logs. This approach allows creating several different instances of input streams with the same distributions.

## VII. JMS MODEL ADEQUACY

The variant of determining the adequacy measure proposed by the authors is based on the proposed in [16] the model's reliability evaluation method — event validity, when comparing event streams of simulated and real systems. In the paper [16], no numerical indicators allowing comparing two event streams are provided.

Let us define the proximity measure of two event streams as follows. We formulate criteria for the unreliability of the predictive model. A model is defined as unreliable if the events number in the simulation did not coincide with the number of events in the real system. If any of the events were not reproduced in the simulation, or new events have arisen, then the model is unreliable. We also consider the model unreliable if the job submit time in the model and the real system do not match, if the job execution time or ordered computing resources do not coincide. Thus, the model is unreliable if  $n \neq N$ ,  $r_i \neq R_i$ ,  $p_i \neq P_i$ , or  $e_i \neq E_i$ .

The number, the order and the time of occurrence of all events are coincided in the experiment and in the real system for a completely reliable model. In practice, the construction of a fully reliable forecasting JMS model is practically impossible even for a natural experiment, as shown above.

Let us consider two model streams of events represented by jobs  $j = (j_1, j_2, \dots, j_n)$  and  $J = (J_1, J_2, \dots, J_N)$ . The job

characteristics  $j_i = r_i$  (submit time),  $p_i$  (resources required),  $e_i$  (required processing time),  $w_i$  (real processing time),  $s_i$  (job launch time). Similar characteristics has the job  $J_i = R_i, P_i, E_i, W_i, S_i$ . The difference measure will be not determined if in the streams do not consider either the number of jobs  $n \neq N$ , or the submit times of any job  $r_i \neq R_i$ , or the ordered resources and processing times for any job  $p_i \neq P_i$ ,  $e_i \neq E_i$ . In this regard, the characteristics can be rewritten as follows:  $J_i = r_i, p_i, e_i, W_i, S_i$ .

Let us construct two vectors of dimension  $n = N$ . For the stream  $j$  we define the vector of job residence times in the system  $v = (v_1, v_2, \dots, v_n)$ ,  $i \in (1, \dots, n)$ , where each component corresponds to the job number in the order in which it enters the system. The value of the component  $v_i = (s_i - r_i + w_i)$  is defined as the residence time of the job in the system, that is, the sum of the wait time and the processing time. For the stream  $J$  we similarly define the vector  $V = (V_1, V_2, \dots, V_n)$ ,  $V_i = (S_i - R_i + W_i)$ ,  $i \in (1, \dots, n)$ .

Thus, we obtained two vectors,  $v$  and  $V$ , the difference between the components of which actually determines the difference between the two JMS models. A natural measure of the proximity of two  $n$ -dimensional vectors is the Euclidean distance between them:

$$E = \sqrt{\sum_{i=1}^n (V_i - v_i)^2} \quad (1)$$

As shown by our experiments, the Euclidean distance increases with the number of processed jobs in the compared experiments. This dependence makes the Euclidean distance inapplicable as a measure of adequacy. We will normalize measure (1) and obtain the measure of the difference  $P$  of the streams  $j$  and  $J$ :

$$P = \sqrt{\frac{\sum_{i=1}^n (V_i - v_i)^2}{n}} \quad (2)$$

The measure of the difference  $P$  (2) does not depends with the number of jobs processed. This fact makes it possible to use the measure  $P$  as a measure of the model adequacy for experiments of any duration.

## VIII. THE REFERENCE VALUE OF THE MEASURE THE JMS MODEL ADEQUACY

The following method is proposed for determining the adequacy measure. The stream  $j$  is determined based on the statistics analysis of a real supercomputer work over a sufficiently long period, and so is the vector  $v$  on stream  $j$  basis. The events  $s_i$  related to the moments of launching jobs (internal scheduler events) are excluded from the stream  $j$ . The selected substream of external events is fed to the JMS model input, and as a simulation result, the stream  $J$  and the corresponding vector  $V$  are generated. The measure  $P$  of the difference between the streams is calculated. The smaller the value of  $P$ , the more adequate the JMS model.

When  $P = 0$ , the JMS model will be completely reliable. The question arises about the maximum permissible value of

the measure  $P_{\max}$ , such that a model with an adequacy measure  $P \leq P_{\max}$  will be considered adequate.

As was shown above, the repetition of a natural experiment does not give a precise reproduction of the result. At the same time, since the real JMS is adequate to itself, some measure  $P_{\text{ideal}}$  of the difference between the streams  $j$  and  $J$ , obtained during two repetitions of the same natural experiment, by definition will be less than the acceptable adequacy limit:  $P_{\text{ideal}} \leq P_{\max}$ .

Let us call  $P_{\text{ideal}}$  the reference value of the adequacy measure. Any model that has an adequacy measure less than or equal to the reference value does not differ in its behaviour from the real system.

Since, for the reasons listed above, carrying out two identical natural experiments in practice is very difficult, it is proposed to determine the adequacy measure reference value by comparing the results of JMS simulation with virtual nodes. We formed a model stream of 1000 jobs based on the statistics of the supercomputer MVS-10P OP installed in the JSCC RAS. This stream was used to model a Russian job management system SUPPZ with virtual nodes. The results are presented in Table 1. The column «number of jobs» corresponds to the number  $k$  of the first jobs of the stream  $j$  (the real SUPPZ) and the stream  $J$  (the SUPPZ with virtual nodes). The column «the number of different jobs» indicates the number of jobs for which the wait times were different in the streams  $j$  and  $J$ . From table 1 we can conclude that the reference value of the JMS model adequacy measure, calculated by the formula (2), is equal to 12.

### III. CONCLUSION

This paper attempts to determine the JMS model adequacy measure by comparing the characteristics of two job streams, one of which is derived from a real supercomputer and the other is derived from the JMS model. Each job in these streams is associated with a set of events – entering the queue, launching, completion. The authors reduced all the events of the job stream into a single vector, in which each component corresponds to a specific job and contains the time that job has spent in the system. The following pairs of vectors are explored in the article: the first vector was acquired from the job streams in the real system and the second one was the generated by JMS model.

TABLE I.  
MEASURES OF JOB STREAMS DIFFERENCE FOR THE SUPPZ WITH  
VIRTUAL NODES

Number of jobs (size of compared vectors)	Measure of stream difference	The number of different jobs
50	0	0
100	12.0	4
250	11.4	13
500	12.0	20
750	11.6	28
1000	11.2	35

It is shown that the normalized Euclidean distance between the vectors in the pair can be used as a JMS model adequacy measure. Besides that, the paper defines the adequacy measure reference value corresponding to the JMS model with virtual nodes.

### REFERENCES

- [1] A. Reuther, et al., "Scalable system scheduling for HPC and big data," in *Journal of Parallel and Distributed Computing*, vol. 111, 2018, pp. 76–92. <https://dx.doi.org/10.1016/j.jpdc.2017.06.009>
- [2] A.B. Yoo, M.A. Jette, M. Grondona, "SLURM: Simple Linux Utility for Resource Management," in *Lecture Notes in Computer Science*, vol 2862, 2003, pp. 44–60. [https://dx.doi.org/10.1007/10968987\\_3](https://dx.doi.org/10.1007/10968987_3)
- [3] R.L. Henderson, "Job scheduling under the Portable Batch System," in *Lecture Notes in Computer Science*, vol 949, 1995, pp. 279-294. [https://dx.doi.org/10.1007/3-540-60153-8\\_34](https://dx.doi.org/10.1007/3-540-60153-8_34)
- [4] SUPPZ. (In Russian) URL: <http://suppz.jssc.ru/> (accessed: 23.04.2019).
- [5] A.V. Baranov, D.S. Lyakhovets, "Comparison of the Quality of Job Scheduling in Workload Management Systems SLURM and SUPPZ," in *Scientific Services & Internet: All Facets of Parallelism: Proceedings of the International Supercomputing Conference*, 2013, pp. 410–414 (in Russian).
- [6] N.A. Simakov et al., "A Slurm Simulator: Implementation and Parametric Analysis," in *Lecture Notes in Computer Science*, vol 10724, 2017, pp. 197-217. [https://dx.doi.org/10.1007/978-3-319-72971-8\\_10](https://dx.doi.org/10.1007/978-3-319-72971-8_10)
- [7] I.M. Yakimov, M.V. Trusfus, V.V. Mokshin, and A.P. Kirpichnikov, "AnyLogic, ExtendSim and Simulink Overview Comparison of Structural and Simulation Modelling Systems," in *Proc. 3rd Russian-Pacific Conference on Computer Technology and Applications (RPC)*, Vladivostok, 2018, pp. 1-5. <https://dx.doi.org/10.1109/RPC.2018.8482152>
- [8] S.W. Cox, "GPSS World: A brief preview," in *1991 Winter Simulation Conference Proceedings*, Phoenix, AZ, USA, 1991, pp. 59-61. <https://dx.doi.org/10.1109/WSC.1991.185591>
- [9] S.R. Chelladurai, "Gridsim: a flexible simulator for grid integration study," 2017. <https://dx.doi.org/10.24124/2017/1375>
- [10] R.N. Calheiros, R. Ranjan, A. Beloglazov, C.A. De Rose, and R. Buyya, "CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms," in *Softw.: Pract. Exper.*, 2011, pp. 23-50. <https://dx.doi.org/10.1002/spe.995>
- [11] W. Chen, and E. Deelman, "WorkflowSim: A toolkit for simulating scientific workflows in distributed environments," in *IEEE 8th International Conference on E-Science, Chicago, IL*, 2012, pp. 1-8. <https://dx.doi.org/10.1109/eScience.2012.6404430>
- [12] H. Xia, H. Dail, H. Casanova, and A.A. Chien, "The MicroGrid: using online simulation to predict application performance in diverse grid network environments," in *Proc. of the 2d Int. Workshop on Challenges of Large Applications in Distributed Environments*, 2004, pp. 52-61. <https://doi.org/10.1109/clade.2004.1309092>
- [13] W. Cirne, and F. Berman, "A model for moldable supercomputer jobs," in *Proc. 15th International Parallel and Distributed Processing Symposium. IPDPS 2001*, San Francisco, CA, USA, 2001, p. 8. <https://dx.doi.org/10.1109/IPDPS.2001.925004>
- [14] Standard Workload Format. URL: <http://www.cs.huji.ac.il/labs/parallel/workload/swf.html> (accessed 24.04.2019)
- [15] U. Lublin, D.G. Feitelson, "The workload on parallel supercomputers: modeling the characteristics of rigid jobs," in *Journal of Parallel and Distributed Computing*, vol. 63, issue 11, 2003, pp 1105-1122. [https://dx.doi.org/10.1016/S0743-7315\(03\)00108-4](https://dx.doi.org/10.1016/S0743-7315(03)00108-4)
- [16] B.M. Glinsky, A.S. Rodionov, M.A. Marchenko, D.I. Podkorytov, and D.V. Weins, "Agent-Oriented Approach to Simulate Exaflop Supercomputer with Application to Distributed Stochastic Simulation," in *Bulletin of the South Ural State University, Series «Mathematical Modelling, Programming & Computer Software»*. 2012, no 18(277), pp. 93-106 (in Russian).

# Towards Big Data Solutions for Industrial Tomography Data Processing

Aleksandra Kowalska<sup>1</sup>, Piotr Łuczak<sup>2</sup>, Dawid Sielski<sup>3</sup>, Tomasz Kowalski<sup>4</sup>,  
 Andrzej Romanowski<sup>5</sup> and Dominik Sankowski<sup>6</sup>

Institute of Applied Computer Science, *Lodz University of Technology*  
 Łódź, Poland

<sup>1</sup>akowalska@kis.p.lodz.pl, <sup>2</sup>pluczak@kis.p.lodz.pl, <sup>3</sup>dawid.sielski@outlook.com, <sup>4</sup>t.kowalski@kis.p.lodz.pl,  
<sup>5</sup>androm@kis.p.lodz.pl, <sup>6</sup>dsan@kis.p.lodz.pl

**Abstract**—This paper presents an overview of what Big Data can bring to the modern industry. Through following the history of contemporary Big Data frameworks the authors observe that the tools available have reached sufficient maturity so as to be usable in an industrial setting. The authors propose the concept of a system for collecting, organising, processing and analysing experimental data obtained from measurements with process tomography. Process tomography is used for noninvasive flow monitoring and data acquisition. The measurement data is collected, stored and processed to identify process regimes and process threats. Further general examples of solutions that aim to take advantage of the existence of such tools are presented as proof of viability of such approach. As the first step in the process of creating the proposed system, a scalable, distributed, containerisation-based cluster has been constructed, with consumer-grade hardware.

**Index Terms**—Big Data, Process Tomography, data processing, data acquisition

## I. INTRODUCTION

MEASUREMENT technologies are a practical challenge for engineers who are increasingly improving them, with novel algorithms for solving optimisation problems being continuously developed.

With the advent of industry 4.0 [1], the volume of the measurement data generated by industrial process becomes too large for processing on a single workstation. Through the use of wireless sensor networks, measurements can be taken in places previously inaccessible for traditional, wired solutions, hence allowing for preventative repair of equipment before the actual failure occurs [2]. As such new, sensor-rich systems are created there exists an unprecedented opportunity to derive deeper insights regarding the nature of the whole process from the large volume of collected data.

An example of a system rich in sensors is process tomography. It is a rapidly developing non-invasive diagnostic technique, finding wider and wider applications in various fields of science. In recent years, process tomography applications have been developed in the petrochemical, pneumatic and gravitational transport of bulk materials, as well as in the pharmaceutical industry and biomedicine. Fig. 1 shows different types of process tomography systems.

In the issues of non-invasive diagnostics and monitoring, various types of tomographic sensors can be combined to provide multi-modality, versatility and adaptation to the dynamics

of the industrial process. The tomographic systems during computer diagnostics of the industrial process can also be supported by additional measurement sources, including ultra-fast video cameras, dedicated flow meters, scales, pressure and temperature sensors. In this way, additional information about the flow and its parameters is obtained. In addition, measurements from tomographic systems can be further used to reconstruct two- and three-dimensional images - both raster sequences and movie sequences. If this data comes from many sensors at the same time, is collected with high time resolution, and the industrial process is long-lasting, a large amount of data appears, which needs to be structured and categorised in appropriate database structures.

The diagnostic information collected using the process tomography measurement systems over a longer period of time can be characterised by the size of even a dozen terabytes, which predestines it for the term Big Data. In particular, such a term can be defined as long-term acquisitions of three-dimensional tomographic images with high temporal-spatial resolution, video recordings, sets of mathematical-physical models and descriptions of experiments. Therefore, this paper presents the concept of a system for collecting, organising, processing and analysing experimental data obtained from measurements using process tomography. The results of conducted research will enable computer systems of non-invasive industrial diagnostics to make quick and reliable decisions in the field of monitoring and control.

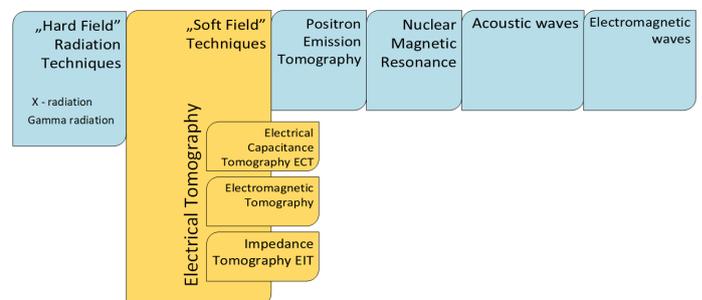


Fig. 1. Types of process tomography measurement modalities. Second from the left illustrates electrical tomography systems.

The contribution of this paper is the proposal for an ar-

chitecture based on containerised instances of Apache Spark and Hadoop which promises to offer soft real-time processing capabilities, whilst being trivially extendable to any number of distributed machines. Since all services in the proposed system are containerised, it is capable of being expanded or shrunk to fit the current needs through the simple process of either starting or shutting down a number of container instances.

## II. BIG DATA IN INDUSTRIAL APPLICATIONS

### A. Evolution of Big Data solutions

The Big Data technology as known today, begun with the creation of the Google File System which enabled large scale data storage while using low-cost commodity hardware [3]. While providing no distributed computational capability, it contributed fault tolerant, multi-access file storage which facilitated concurrent work on the same file, hence implementing distributed producer-consumer queues for large (Multi-GB) files. This architecture, albeit limited, provided the foundation for the systems capable of handling both high volume data at rest (static files) and in motion (dynamically appended files), thus inherently satisfying both the Volume and Velocity part of the five V's of Big Data [4].

The next step in the development of a full Big Data stack was the MapReduce paradigm [5]. This paradigm was created as a means of unifying the disparate solutions that previously had to split their focus between the problems of pluralisation and the actual problem hence obscuring the latter. This new paradigm was a reaction to this unwanted, additional complexity and provided a simple abstraction layer allowing the programmers to specify their computations in the form of a chain of *map* and *reduce* steps, hence its name. The *map* function is responsible for processing the input key-value pair into intermediate results, also in the form of a key-value pair, whilst the *reduce* step is responsible for merging the intermediate values into the final result. This system, while capable of operating separately from the distributed file system, generally exists in symbiosis with the storage nodes of such a file system, thus resulting in the popular technique used to remove bottlenecks in the data processing, which generally referred to as "bringing processing to the data". The combined GFS and MapReduce provided the foundation for modern Big Data processing systems capable of satisfying all five V's, namely the aforementioned Volume and Velocity, and the previously unmentioned Veracity, Variety and Value.

Whilst the initial advancements in distributed file storage were proprietary and generally only available in the companies on the bleeding edge of Big Data technology, the Hadoop Distributed File System alongside with the rest of the Hadoop project were created by the Apache foundation, thus resulting in a significantly lowered barriers to entry into the Big Data industry [6]. The Open Source nature of this project also provided a good basis for the development of new data processing tools hence becoming an unofficial standard for big data storage and analysis systems.

One of the limitation of the early MapReduce systems was their inherent assumption that the data flow is acyclic, that is

that a computational task begins with data being read from the file system and ends with the resulting data being written back to said file system. This, combined with the fact that distributed file systems, such as the GFS and HDFS were optimised for high throughput and not low file access latency, meant that, while possible, the execution of tasks that reused the working set of data was inadvertently slowed down by the file systems access time. In order to facilitate the execution of such tasks, a new framework named Spark was proposed [7]. Spark's Resilient Distributed Datasets (RDDs) provided a way for the data to be cached in memory during the execution of the iterative task, hence avoiding the performance penalty resulting from repeated file system accesses. This novel approach allowed for a significant decrease in the execution time of iterative tasks such as linear regression or the alternating least squares computation, making Big Data based machine learning solutions viable.

With the proliferation of new data processing frameworks the tight coupling between the MapReduce programming model and the Hadoop file system became an impediment which had to be worked around by the developers of these solutions. A new incarnation of the Hadoop framework was created with the aim of alleviating the aforementioned limitations and clearly separating the resource management from the programming model. The framework was named Yet Another Resource Negotiator (YARN) [8]. In this framework, MapReduce was no longer the primary focus and thus became simply one of the possible tools that could be run. This new version of MapReduce is commonly referred to as MapReduceV2. A multitude of programming frameworks, that initially had to be built on top of MapReduce and thus were beholden to its limitations, were updated to use this new model even before YARN left the beta stage of its development, therefore validating the design decisions made by the YARN creators.

As a result of the aforementioned progress the Big Data ecosystem became not only a feature rich set of tools but also a stable and mature foundation for building new solutions and applications that deal with large volumes of data [9].

### B. Industrial applications

Modern industry tends to generate an enormous amount of data every day while in many cases lacking the technical capabilities required effectively process and derive long-term insights from it [10] [11]. In many cases the extent of use of this data boils down to being shown to the operator on-site so as to facilitate the monitoring of the industrial process. Such analysis is not only limited by the finite capabilities of both human mind and memory but also results in very low bus factor, making the expert an unexpedient component of the process. This situation however is slowly changing with new approaches and initiatives such as the industry 4.0.

The financial industry has devoted considerable research efforts into what could be considered proto-Big Data processing since as early as 1997 [12]. In this sector one of the most important computational tasks is fraud detection, which unlike most statistical data analysis, concerns itself with

data in motion, with real time emphasis. Some AI application require enormous collections of data to be stored, prepared and processed by machine learning such as the EEG annotation data [13], industrial emergency states detection [14] [15] or medical augmented reality future diagnosis [16].

### III. STATE OF TECHNOLOGY

As presented in II-B there exist numerous applications of big data in different domains. One of the interesting examples was reported by Skuza et al. [17], another example of the barely explored domain is the usage of big data for process tomography [18]. Process tomography is a set of techniques that are responsible for acquiring data from a given process, processing them and providing information about concentration distribution and flow.

Process tomography together with big data solves lots of different problems that emerge with traditional approach. Romanowski et al. started to join these two fields to propose a solution for detecting material plugs in pneumatic conveying measurement data [18] [19]. The authors used Hadoop (installed on three independent computers) together with Mahout machine learning library. The data consisted of five experiments performed on several different pipe diameters with different flow conditions. The tests were performed in two distinct locations at different times. Total data obtained during the data acquisition part was about 40 GB but it must be stated that the computational system is able to handle significantly larger amount of data. The authors present proof that the combination of big data tools and conventional algorithms can be used in automatic detection of material plugs in vertical or horizontal flows. This indicates that using big data tools such as Hadoop is an excellent alternative for traditional approach.

Process tomography uses lots of different techniques which include lots of different sensors from which the data is obtained. Numerous different sensors with numerous output format are a problem designed to be solved by big data. Rymarczyk et al. propose and discuss the system which is designed to optimize and automate given process by using numerous tomographic sensors [20]. The paper indicates the importance of using such systems for maintaining competitiveness. The design of the system includes multiple sensors like ECT (Electrical Capacitance Tomography), ERT (Electrical Resistivity Tomography), UST (UltraSound Tomography), together with temperature and pressure information. In the next step the authors perform the data acquisition during which the data is saved to a server for later processing. Next part uses image reconstruction together with cloud computing which later is employed to control industrial processes.

There are numerous steps needed to be performed in the system in order to obtain data required to control some process. There is also an increase in amount of data that needs to be processed. As a result the data processing part may take a longer amount of time. To lower that the parallelisation can be introduced. This approach was used by Chen et al. where he designed and tested solution that uses crowd-sourcing to help with understanding the particle-tracking problem [21]. The

authors addressed that at the time being there are no solutions that can accurately analyse flow in silos. This resulted in building a system that employed experts and non-experts to analyse tomographic images. The results from this system were compared to the results from the automatic approach. The paper clearly states that the crowd solution is significantly better in terms of scale, delivered result and economics in comparison to the automatic approach. What is more, the system can be applied in different domains which makes it generalisable.

The goal is to produce system that will be responsible for controlling industrial process in order not only to prevent undesirable behaviour from happening but also to obtain better understanding about the process along with the possibility of optimisation of the process. The newest approach proposed by Romanowski gives more insight on joining big data with process tomography by extending work presented by Chen and is an example of the desired system [4]. The author in his work designed a methodology that improves current state of knowledge, regarding recognition of specific bulk flow regimes (e.g. pipeline blockage threats). To obtain such improvement the author presented solution using Hadoop platform for distributed data processing on large quantities of data using cluster computing. For the classification part the Apache Spark with SVM as an algorithm was used as a supervised learning method. The system can also be employed to handle processes from different domains. The main advantage of this solution is the improvement in finding flow regimes that may be critical for the process and perform actions to prevent such flow. This work also describes how to jointly analyse incoherent data.

### IV. GOALS, VISION AND PROPOSED APPROACH

The aim of the work is to develop new algorithms for the collection, organisation, processing and analysis of large data sets obtained from industrial systems of non-invasive diagnostics and control, based on process tomography.

As part of the work, it is proposed to use a containerisation-based [22] implementation system that will allow for easy increase of available disk space and computing power, both using an external cloud infrastructure and generally available consumer computer equipment. Due to the particular requirements of the platform components, it may prove inadvisable for such an approach to cover the entire project. The use of containerisation has certain consequences and limitations that should be considered. Containers introduce another (though lightweight) layer "reflecting" a given element of the platform from the concept of the real-time system. Stations and diversified installation software have specific hardware requirements (eg. control interfaces) or visualisation (eg. the use of calculations directly from the GPU), which may potentially disqualify the software as a service subject to containerisation. In connection with the above, work on the implementation system based on the above approach has been focused on the computational cluster responsible for Big data analysis of historical data and eventually, real-time sensor readings. Fig. 2 presents the proposed model in which the

dependence of the control loop on the cloud is limited to "fine tuning" so as to avoid affecting the delay of regulation.

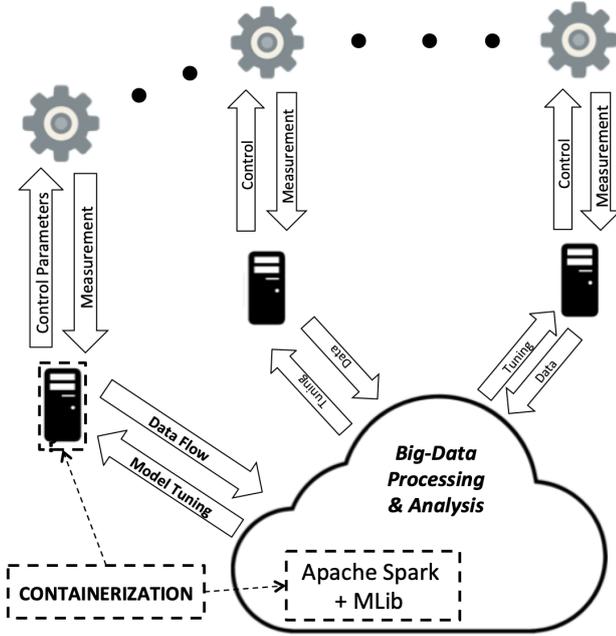


Fig. 2. Proposed model

Based on the aforementioned attempts, from section III, at creating data based systems for the purposes of analysis and control of tomographic processes, the planned focus of our research is twofold. Firstly, we plan to assess the benefit, defined as an increase in the computation speed of visualisation of the measured flow, of using a cluster consisting of consumer-grade hardware over the traditional approach of carrying out the computations on one high-end workstation. Secondly, we plan to assess the viability of using such cluster as a perpetually self-improving industrial controller. It is our belief that, provided enough training data, such a system could potentially operate with only minimal supervision from an operator, hence allowing a single expert to oversee a larger number of apparatus than it is currently possible. This concept will require an extensive amount of experimentation before it can be considered ready for industrial use, though at this early stage two criteria of viability can be defined:

- 1) capability of gradual auto-tuning of traditional controller
- 2) nearly real-time performance of the flow analysis

Should only one of these criteria be met the system would still be a valuable enhancement of a traditional system, either as an "auto-tuner" or as a real-time (or almost real-time), flow visualisation platform.

Should both of these criteria be met the future steps would include a "run-ahead", real-time simulation that could be used to predict potential faults and act accordingly to prevent them. An alternative approach for this purpose is also considered, namely instead of simulating the flow, an additional machine learning model could be used as a fault predictor. The architecture of the proposed cluster can be seen on figure 3.

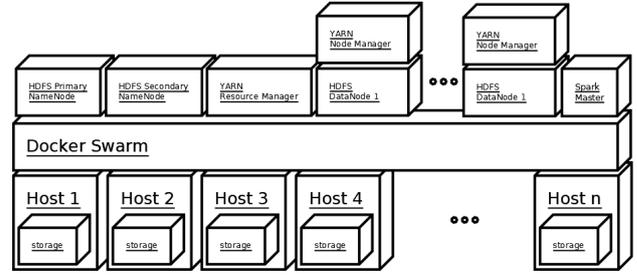


Fig. 3. Structure of the proposed cluster

## V. EVALUATION

The initial evaluation of the system generated promising results, as the even with as few as 5 low cost machines it was capable of outperforming the existing single-machine solution present in the Tom Dyakowski Process Tomography Laboratory in Lodz University of Technology, shown on figure 4. In the initial testing, which used a very limited number of machines with the exact same hardware, the increase in the speed of computations (measured as a decrease in computation time) was linear, whilst the latency did not exceed 2.3 seconds. Due to a very limited number of machines with the exact same specifications, subsequent tests had to involve a variety of configurations. Since the machines available did not have the



Fig. 4. Tom Dyakowski Process Tomography Laboratory: experimental silo (left) and data acquisition, processing and visualisation workstation (right)

exact same hardware specification, it is impossible to measure a precise number of machines at which the cluster becomes a more efficient computational environment, however this non-uniformity proves to be a notable advantage of our approach. Since the cluster was capable of working with machines with differing specifications, one could construct such a solution without incurring additional cost, by simply using the spare, possibly outdated, hardware already available on-site.

## VI. DISCUSSION AND FUTURE WORK

This paper presents the idea of a big data system based on tomographic solutions. The general structure of the cluster is shown in Figure 3. Particularly noteworthy here is the ease of expanding the system described, both in terms of

physical devices and additional services. An example of such an extension is the planned addition of more efficient nodes equipped with an increased amount of volatile memory and a high-end graphics card in order to facilitate parallel computing.

As part of further work, proprietary algorithms for automatic acquisition, analysis and interpretation of large data sets [23] will be developed, as well as intelligent decision algorithms supporting effective diagnostics and monitoring of flow processes [14]; including those to be the next step after crowdsourcing data labelling cases impossible to be automatically processed at once [24] [25]. The prepared algorithms will be verified both by simulation and by real experimental data from their own experiments carried out on the basis of unique, semi-industrial research installations. The results of research will enable computer systems to make quick decisions in the field of monitoring and control. Furthermore, the computational environment may be suited to personal-medical applications, such as the continuous glucose monitoring, resulting in better diagnosis, safety and hence better life comfort [26].

## VII. CONCLUSION

This work presents a pioneering concept for the Big Data system. The solution concerns the integration of data from sensors used in process tomography, which enables imaging and feedback associated with control in the field of process optimisation. The presented infrastructure can bring tangible benefits in various sectors of the industry due to its scalable nature, allowing for smooth expansion as the company or its requirements grow. The described approach is based on a Docker Swarm cluster which facilitates easy fail-over in case of hardware node failures.

## ACKNOWLEDGMENT

This work is partially financed by the Smart Growth Operational Programme 2014–2020 project no POIR.04.01.02-00-0089/17-00. The project is conducted in the Institute of Applied Computer Science at the Lodz University of Technology.

## REFERENCES

- [1] H. Lasi, P. Fettke, H.-G. Kemper, T. Feld, and M. Hoffmann, "Industry 4.0," *Business & Information Systems Engineering*, vol. 6, no. 4, pp. 239–242, 2014. doi: 10.1007/s12599-014-0334-4
- [2] V. C. Gungor, G. P. Hancke *et al.*, "Industrial wireless sensor networks: Challenges, design principles, and technical approaches." *IEEE Trans. Industrial Electronics*, vol. 56, no. 10, pp. 4258–4265, 2009. doi: 10.1109/TIE.2009.2015754
- [3] S. Ghemawat, H. Gobioff, and S.-T. Leung, "The Google file system," p. 29, 2003.
- [4] A. Romanowski, "Big data-driven contextual processing methods for electrical capacitance tomography," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 3, pp. 1609–1618, 2019. doi: 10.1109/TII.2018.2855200
- [5] J. Dean and S. Ghemawat, "Mapreduce: simplified data processing on large clusters," *Communications of the ACM*, vol. 51, no. 1, pp. 107–113, 2008. doi: 10.1145/1327452.1327492
- [6] K. Shvachko, H. Kuang, S. Radia, and R. Chansler, "The hadoop distributed file system," in *Mass storage systems and technologies*, 2010. doi: 10.1109/MSST.2010.5496972 pp. 1–10.
- [7] M. Zaharia, M. Chowdhury, M. J. Franklin, S. Shenker, and I. Stoica, "Spark: Cluster computing with working sets." *HotCloud*, vol. 10, no. 10-10, p. 95, 2010.
- [8] V. K. Vavilapalli, A. C. Murthy, C. Douglas, S. Agarwal, M. Konar, R. Evans, T. Graves, J. Lowe, H. Shah, S. Seth *et al.*, "Apache hadoop yarn: Yet another resource negotiator," in *Proceedings of the 4th annual Symposium on Cloud Computing*, 2013. doi: 10.1145/2523616.2523633
- [9] P. Basanta-Val, N. C. Audsley, A. J. Wellings, I. Gray, and N. Fernández-García, "Architecting time-critical big-data systems." *IEEE Transactions on Big Data*, vol. 2, no. 4, pp. 310–324, 2016. doi: 10.1109/TB-DATA.2016.2622719
- [10] K. Grudzien, A. Romanowski, D. Sankowski, and R. A. Williams, "Gravitational granular flow dynamics study based on tomographic data processing," *Particulate Science and Technology*, vol. 26, no. 1, pp. 67–82, 2007. doi: 10.1080/02726350701759373
- [11] T. Rymarczyk, "Using electrical impedance tomography to monitoring flood banks," *International Journal of Applied Electromagnetics and Mechanics*, vol. 45, pp. 489–494, 2014. doi: 10.3233/JAE-141868
- [12] K. Grudzien, A. Romanowski, and R. A. Williams, "Application of a bayesian approach to the tomographic analysis of hopper flow," *Particle & Particle Systems Characterization*, vol. 22, no. 4, pp. 246–253, 2005. doi: 10.1002/ppsc.200500951
- [13] S. Opałka, B. Stasiak, D. Szajerman, and A. Wojciechowski, "Multi-channel convolutional neural networks architecture feeding for effective eeg mental tasks classification," *Sensors*, vol. 18, no. 10, 2018. doi: 10.3390/s18103451
- [14] A. Romanowski, "Contextual processing of electrical capacitance tomography measurement data for temporal modeling of pneumatic conveying process," in *Proceedings of the 2018 Federated Conference on Computer Science and Information Systems*, vol. 15, 2018. doi: 10.15439/2018F171 pp. 283–286.
- [15] K. Grudzien, A. Romanowski, and R. A. Williams, "Application of a bayesian approach to the tomographic analysis of hopper flow," *Particle & Particle Systems Characterization*, vol. 22, no. 4, pp. 246–253, 2005. doi: 10.1002/ppsc.200500951
- [16] A. Nowak, M. Wozniak, M. Pieprzowski, and A. Romanowski, "Towards amblyopia therapy using mixed reality technology," in *2018 Federated Conference on Computer Science and Information Systems (FedCSIS)*, 2018. doi: 10.15439/2018F335 pp. 279–282.
- [17] M. Skuza and A. Romanowski, "Sentiment analysis of twitter data within big data distributed environment for stock prediction," in *2015 Federated Conference on Computer Science and Information Systems (FedCSIS)*, 2015. doi: 10.15439/2015F230 pp. 1349–1354.
- [18] A. Romanowski, M. Skuza, P. Wozniak, K. Grudzien, and Z. Chaniecki, "Big data computational environment for tomography measurement data," *Process Tomography WCIPT7, Poland*, 2013.
- [19] A. Romanowski, K. Grudzien, Z. Chaniecki, and P. Wozniak, "Contextual processing of ECT measurement information towards detection of process emergency states," in *13th International Conference on Hybrid Intelligent Systems*, 2013. doi: 10.1109/HIS.2013.6920448 pp. 291–297.
- [20] T. Rymarczyk and J. Sikora, "Applying industrial tomography to control and optimization flow systems," *Open Physics*, vol. 16, p. 46, 2018. doi: 10.1515/phys-2018-0046
- [21] C. Chen, P. W. Woźniak, A. Romanowski, M. Obaid, T. Jaworski, J. Kucharski, K. Grudzień, S. Zhao, and M. Fjeld, "Using crowdsourcing for scientific analysis of industrial tomographic images," *ACM Trans. Intell. Syst. Technol.*, vol. 7, no. 4, p. 52, 2016. doi: 10.1145/2897370
- [22] W. Felter, A. Ferreira, R. Rajamony, and J. Rubio, "An updated performance comparison of virtual machines and linux containers," in *2015 IEEE International Symposium on Performance Analysis of Systems and Software*, 2015. doi: 10.1109/ISPASS.2015.7095802 pp. 171–172.
- [23] A. Kowalska, R. Banasiak, A. Romanowski, and D. Sankowski, "3d-printed multilayer sensor structure for electrical capacitance tomography," *Sensors*, vol. 19, no. 15, 2019. doi: 10.3390/s19153416
- [24] A. Romanowski, P. Łuczak, and K. Grudzień, "X-ray imaging analysis of silo flow parameters based on trace particles using targeted crowdsourcing," *Sensors*, vol. 19, no. 15, 2019. doi: 10.3390/s19153317
- [25] I. Jelliti, A. Romanowski, and K. Grudzień, "Design of crowdsourcing system for analysis of gravitational flow using x-ray visualization," in *Proceedings of the 2016 Federated Conference on Computer Science and Information Systems*, M. Ganzha, L. Maciaszek, and M. Paprzycki, Eds., vol. 8, 2016. doi: 10.15439/2016F288 pp. 1613–1619.
- [26] P. Kucharski, K. Pagacz, A. Szadkowska, W. Młynarski, A. Romanowski, and W. Fendler, "Resistance to data loss of glycemic variability measurements in long-term continuous glucose monitoring," *Diabetes Technology & Therapeutics*, vol. 20, no. 12, pp. 833–842, 2018. doi: 10.1089/dia.2018.0247



# Whose Fault is It?

## Correctly Attributing Outages in Cloud Services

Matteo Adriani

*Dpt. of Civil Engineering and Computer Science*  
*University of Rome Tor Vergata*  
Rome, Italy  
matteo.adriani93@alice.it

Maurizio Naldi

*Dpt. of Civil Engineering and Computer Science*  
*University of Rome Tor Vergata*  
Rome, Italy  
*Dpt. of Law, Economics, Politics and Modern Languages*  
*LUMSA University*  
maurizio.naldi@uniroma2.it  
m.naldi@lumsa.it

**Abstract**—Cloud availability is a major performance parameter in cloud Service Level Agreements (SLA). Its correct evaluation is essential to SLA enforcement and possible litigation issues. Current methods fail to correctly identify the fault location, since they include the network contribution. We propose a procedure to identify the failures actually due to the cloud itself and provide a correct cloud availability measure. The procedure employs tools that are freely available, i.e. traceroute and whois, and arrives at the availability measure by first identifying the boundaries of the cloud. We evaluate our procedure by testing it on three major cloud providers: Google Cloud, Amazon AWS, and Rackspace. The results show that the procedure arrives at a correct identification in 95% of cases. The cloud availability obtained in the test after correct identification lies between 3 and 4 nines for the three platforms under test.

### I. INTRODUCTION

Availability is a major Quality of Service descriptor in cloud services, and an essential component of Service Level Agreements [1]–[3].

Many efforts have been devoted to understanding and improving the availability of cloud systems. The relevance of the issue has been re-stated very recently by Varghese and Buyya, which list it among the top research directions, mentioning the 49-minute outage suffered by Amazon, which cost the company more than \$4 million in lost sales, as an indicator of the economic importance of achieving a high availability [4]. The same concept had been voiced in [5], where the authors even propose to consider a *Reliability as a Service*, where reliability is a parameter that users can specify and a service by itself, rather than the random state of a cloud-based service. Concerns for the legal implications that may arise due to a less-than-adequate cloud reliability have been recently expressed in [6].

An analysis of the main causes of cloud failures has been carried out in [7], where growth trends are also identified, and [8], where mechanisms are subsequently discussed to minimize the impact of outages. Some papers have focussed on the analysis of the cloud architecture to get a high availability by design [9]–[11]. A different approach has been taken in [12] and [13], where machine learning technique have been employed to predict cloud outages (and react accordingly).

If we switch from the perspective of a cloud designer to that of a cloud user, the main interest lies in understanding if the cloud is performing up to the expectations. Setting up, or employing the services of, a cloud monitoring platform is essential in this respect. Several architectures have been proposed for that purpose, e.g. in [14]–[16], and a recent review is contained in [17].

Unfortunately, very few attempts have been done to actually measure cloud availability from a third party vantage point. An early attempt based on users' reports has been reported in [18]. The shortcoming of that approach is that the starting time of the outage may not be reported correctly, since a time lag is always present between the time an outage occurs and the time a user first reports it. The ending time of the outage may be also reported wrongly, since most users do not take on themselves to report it, and we have to rely on the cloud provider announcing that the problem has been solved and the cloud is back to its fully operational state. Statistics of working periods and outages have been modelled in [19] with data coming from a small private cloud. Active measurement systems based on ICMP probing packets have been investigated in [20]–[22]. A major issue with all measurements campaigns conducted so far is that they do measure the quality of service experienced by the user, but in doing so they include the loss contribution provided by the network located between the cloud user and the cloud server. The availability that is measured in the end is an underestimation of the actual cloud availability.

In this paper, we propose a measurement method that allows to distinguish between the losses due to the network and those due to the cloud, returning the true cloud availability. After describing the intrusive network problem in Section II and recalling the definition of availability in Section III, our study provides the following original contributions:

- we propose a measurement procedure to measure true cloud availability (Section IV);
- we assess its success rate (Section V), showing that it outperforms previous methods usable for that purpose;
- we apply our procedure to three major cloud providers and contrast the results with concerns arisen in early measurement campaigns (Section V), showing that the

availability at IP level is close to four nines, and the the network contribution is so relatively large as to significantly alter the overall results in the absence of a correct failure attribution procedure.

## II. THE LONG ROAD TO THE CLOUD

Cloud availability measurements are a major tool in assessing a cloud provider's compliance with SLA targets and obligations. However, those measurements may lead to false conclusions if they are not carried out properly. In this section, we take a look at what is probably the most important reason for lack of accuracy.

Contents placed on a cloud are located among one or more data centers, whose location is, by definition, unknown to the user [23], [24]. Whatever the way by which we probe the cloud to measure availability, third-party measurements are conducted from outside the cloud, i.e. through the network. In probing the cloud, we can therefore mimic the experience of the user, traversing one or more Internet Service Providers (ISP) and several Autonomous Systems (AS), as shown in Fig. 1.

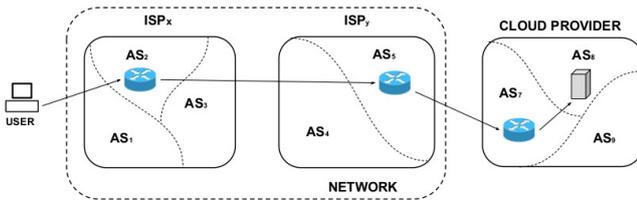


Fig. 1: The path from user to cloud provider.

It has been noted that false outages may be declared, since the lack of response to a user's request to a cloud may be due to packet losses in the network rather than the cloud itself [20]. This appears to be a major problem if we wish to get an accurate measurement of the actual outage rate for the cloud. When measurements are conducted through ICMP probing packets (pings), the Majority Voting rule to declare an outage has been analysed as an effective remedy in several contexts [21]. Under Majority Voting, an outage is declared if a majority of pings get no echoes. However, it cannot be considered as the definitive solution, since its accuracy depends on the specific combination of cloud and network performances.

We therefore need a more generally reliable approach to obtain an accurate measurement of cloud outage in the face of the losses of probing packets due to the network.

This is particularly relevant, since such measurements can be employed to enforce the contractual obligations contained in the SLA and the legal dispute that may arise, a danger that has been dreaded in [6]. Actually, the liability of the cloud provider in the case of obligations related to service malfunctioning has been mentioned as a major obstacle to the wide adoption of the cloud by banks [25]; the same has been reported for the semiconductor industry [26]. If we fail to recognize that service outages may be due to

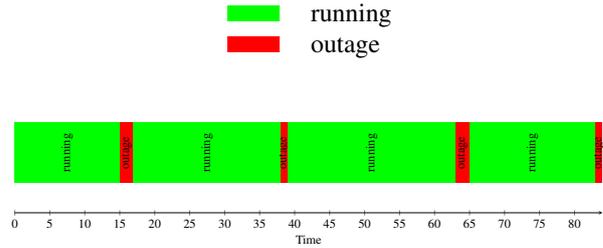


Fig. 2: Cloud state sequence

the network rather than the cloud, the cloud availability is actually underestimated, and the cloud provider may incur undue penalties. At any rate, the overall cost of data center outages is made of many components, which can build up to a very large amount, as reported in a study by the Ponemon Institute [27].

## III. AVAILABILITY OPERATIONAL DEFINITION

Before dealing with the contributions of the network and the cloud to the availability as seen from an external observer, we have to define how the observed availability is measured. In this section, we arrive at the operational definition of availability we have employed in this paper.

For our purposes, the state of the cloud is considered as a succession of working periods and outages, as shown in Fig. 2. If we describe the state of the service through the function  $a(t) : t \rightarrow \{0, 1\}$ , the availability over an observation period  $T$  is then

$$A = \frac{1}{T} \int_0^T a(t) dt. \quad (1)$$

Within this paper, we do not consider the case of graceful degradation, where the cloud service is still running, but with a significantly worsened quality of service. Even though a service may experience a graceful degradation, we imagine that we can always classify the service as either being available or not. For example, if we tolerate a latency lower than a prescribed value, the service may degrade down to that value, while still being considered as available, but will be considered as unavailable when the latency exceeds that threshold.

If we indicate by  $W$  the overall sum of the durations of working periods and by  $F$  the overall sum of the durations of outages, we have the usual definition of availability as the fraction of ON periods over the observation window  $T$

$$A = \frac{W}{W + F}. \quad (2)$$

However, the actual measurement process does not allow to recover the function  $a(t)$ , but rather its sampling version, obtained by probing the system at a discrete set of times. The discrete times are those at which discrete events take place, such as failed or successful service queries.

As a consequence, for cloud services, two general models have been defined in [28] to describe availability from discrete events:

- The dual state model;

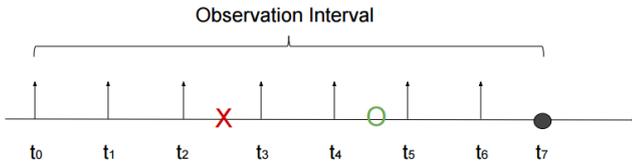


Fig. 3: Probing sequence example

- The success ratio model.

In the dual state model, the availability is computed as a function of the sum of the durations of all down states experienced during contracted service time. In the success ratio model, we refer instead to the event themselves, rather than their duration: the availability is computed as a function of the number of successful and failed resources requests during the contracted service period.

However, if we probe the cloud at periodic intervals (as opposed to random or irregular ones), the distinction between the two models blurs. Considering, e.g., the sequence of probing queries shown in Fig. 3, if we define the down duration as the time distance between the first failed probing query and the first subsequent successful probing query, the two models provide exactly the same availability output (5/7 in this case).

#### IV. CLOUD AVAILABILITY MEASUREMENT

The third-party measurements reported so far in the literature adopt a probing mechanism employing the ping command, which however does not allow to distinguish between outages due to the network and those due to the cloud. In this section, we propose a procedure that allows to obtain the availability of the cloud only. In Section IV-A, we first outline the problems affecting the measurement schemes employed so far, then provide an overview of our new procedure in Section IV-B, and finally describe its phases.

##### A. End-to-end availability

Current procedures to measure the availability of a cloud rely on the use of probing packets sent out from one or several vantage points mimicking the location of a real user. These packets are sent out periodically, as shown in Fig. 3. So far, the ping utility has been used for this purpose. Ping operates by sending Internet Control Message Protocol (ICMP) echo request packets to the target host and waiting for an ICMP echo reply, as shown in Fig. 4 (ping operations are described in Chapter 8.4 of [29]). Echoes from ping are counted as indicators of an operating cloud, while missed echoes are counted as indicators of a failing cloud. It is assumed that a cloud server returning the probing packets is also working correctly to provide services to its clients, i.e., we do not consider software problems related to service provisioning. The ratio of returned echoes to the overall number of sent probes gives us the availability of the cloud.

However, the use of this utility suffers from two main drawbacks:

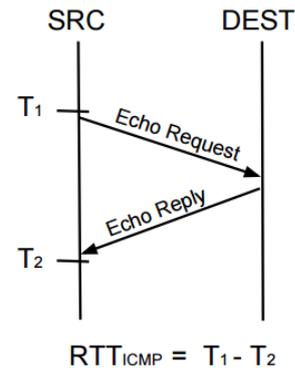


Fig. 4: Ping utility

- it returns an end-to-end measurement that incorporates all the failures taking place on the road to the cloud;
- it employs the ICMP protocol, which may be dealt with differently than TCP/UDP segments, and typically with a lower priority so that the reported availability may be lower than that actually experienced with the cloud-based service.

While the former problem cannot be solved by acting on the probing mechanism alone, the latter problem can be eliminated by employing nping probing packets instead. Nping is an open source tool for network packet generation, response analysis and response time measurement (see Chapter 18 of [30]); it can generate network packets for a wide range of protocols, allowing users full control over protocol headers. We can therefore employ it to generate TCP-like probing packets, which undergo the same priority treatment as the true packets we would employ when using the cloud service <sup>1</sup>.

##### B. Overall procedure

For the time being we consider the reliability at the IP level only, meaning that we are interested in assessing if IP packets transporting the payload involved in the cloud service actually make it through the cloud once they reach it. Our procedure to measure the availability of the cloud, and the cloud only, goes through the following steps:

- 1) Probing the whole sequence of hops along the path from the measurement vantage point to the cloud;
- 2) Associating an ISP to each hop along that path;
- 3) Identifying the first hop belonging to the cloud provider, i.e. the hop marking the entry into the cloud providers domain;
- 4) Counting missing echoes from that first cloud hop and computing the corresponding cloud availability.

In the following, we take care of step 1 in Section IV-C, steps 2 and 3 in Section IV-D, and step 4 in Section IV-E.

##### C. Tracing probing packets

As just recalled, ping (or nping for that purpose) is an end-to-end tool, which does not reveal anything about what

<sup>1</sup><https://nmap.org/nping/>

happens in between the probing source and the end host. We wish instead to get the sequence of IP addresses of routers that make the path from source (our probing vantage point) to destination (the cloud server).

In order to get a complete view of the path from source to destination, along which packets enter the cloud, we can employ the `traceroute` programme<sup>2</sup>. This programme uses limited Time-To-Live (TTL) ICMP probes to discover the IP addresses of IP router interfaces along the path from source to destination, using ICMP echo requests (see Fig. 5). Despite being the most used method to get information about Internet topology, `traceroute` suffers from the following major problems, which may lead to no return from the probed routers or to returned invalid IP addresses:

- ICMP packets may be filtered out by firewalls along the way.
- load-balancing routers may alter the path [31];
- successive TTL-limited packets do not necessarily follow the same forwarding path, so that we may get different chains of routers while we try to discover a single path to destination;
- some hops do not return ICMP replies;
- some routers may be anonymous, i.e., their existence is detected but their interface address is not returned [32];
- some routers may return the address of the interface from which the message came [33];
- some routers return a fixed IP address, regardless of the address of the actual interface on which the message has landed;
- some routers may return an IP address chosen randomly among those of the router's several interfaces;
- the connectivity between routers may be provided through chains of ATM (Asynchronous Transfer Mode) switches or MPLS (MultiProtocol Label Switching) tunnels (reported to account even for 30% of paths [34] [35]), which may make the path opaque to IP probes.

Though the original version of `traceroute` employs ICMP packets, other versions may employ UDP or TCP probing packets.

The UDP version employs limited TTL packets and large destination port numbers. When an intermediate router receives such a probing packet with a zero TTL, it returns an *ICMP time exceeded* message. The source can progressively increase the TTL discovering farther routers along the path, till it reaches the destination. However, the use of UDP messages to high ports shares the same problem with firewall as ICMP packets [36].

A version called `tcptraceroute` has been proposed<sup>3</sup>. The TCP version of `traceroute` bypasses firewalls by directing TCP packets to well-known ports (e.g. port 80), though some firewall may still block TCP packets when no host behind the firewall accepts the TCP connection.

<sup>2</sup><https://wiki.geant.org/display/public/EK/VanJacobsonTraceroute>

<sup>3</sup><https://www.freebsd.org/cgi/man.cgi?query=tcptraceroute&manpath=FreeBSD+9.3-RELEASE+and+Ports>

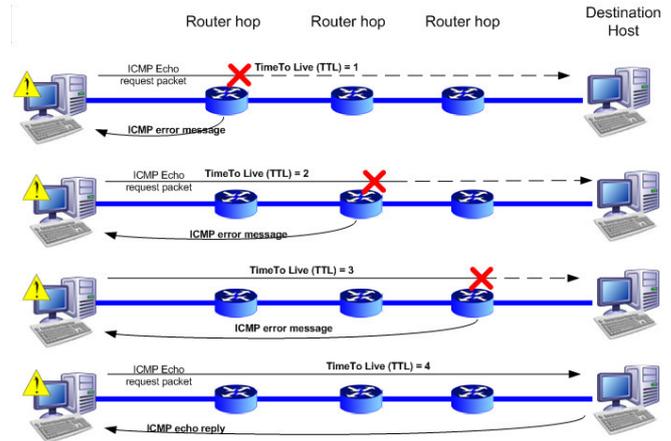


Fig. 5: Working of the `traceroute` programme

Despite the shortcomings of ICMP-based `traceroute`, it appears however to reach targets more successfully than its UDP and TCP counterparts [37]. In this paper, we therefore stick to the classical ICMP-based `traceroute`.

Other methods have been proposed to bypass the limitations of `traceroute`; a recent survey is contained in [38]. The most relevant category at the interface level (which is the one we adopt here) is based on the options of the IP packet header. However, in most cases, it relies either on the cooperation of intermediate routers or on the use of multiple vantage points, which excludes them from the horizon of third-party measurements. In addition, their use increases the chances of packets being discarded or triggering alarms on IDS (Intrusion Detection systems) [39].

#### D. Identifying cloud boundaries

After identifying the chain of routers that lead to the ultimate cloud destination, we wish to identify the AS (Autonomous System) to which each router belongs, and eventually the ISP administering that AS. This is essential to identify the router marking the cloud ingress. Our procedure will go through the following steps:

- 1) Get the ASN associated to each hop on the path to the cloud;
- 2) Get the ISP administering the AS along the path;
- 3) Extract the router marking the ingress into the cloud provider's domain.

Our procedure employs the following protocols and programs:

- `Traceroute`;
- `whois`;
- `RISwhois`.

While the first two are pretty standard programmes, the third one is actually a modified version of `whois`. While the standard version of that protocol queries the Internet Routing registries, the modified version `RISwhois` has been devised in the context

of the RIPE RIS (Routing Information System) project<sup>4</sup> and allows to get data directly from a network of BGP collectors, which collect data from the BGP tables of their peers. Such a use of BGP is envisaged and described in several papers concerned with the need to build the AS topology [38], [40]. Our aim in combining the response from the tools listed above is to arrive at a consensus about the correct ASN to attribute to each hop in the path to the cloud. Our procedure will return a positive result if we achieve either a 2/2 result or a 2/3 results, i.e. at least 2 of the tools agree.

It is to be noted that the `brdmap` programme is also available to identify domain borders, as described in [41], where it has been employed, however, through the use of 19 vantage points, whereas our procedure employs a single vantage point.

The first tool we employ to carry out the IP-to-AS attribution is `traceroute`. For each hop, the `-a` option allows to get the AS number. However, as pointed out in [40], the `traceroute` command alone does not give an accurate AS number in all cases, and it does not even return an AS number in roughly 10% of cases. We consider separately the case where `Traceroute` returns an ASN for each hop and that where it does not.

If `Traceroute` does not return an ASN number, we resort to `whois` and `RISwhois` in parallel. If they both return the same ASN, then we consider that to be the correct ASN; otherwise (i.e., if they either return different ASNs or do not return an ASN at all), the procedure is considered to fail.

If `Traceroute` does return an ASN, however, we do not stay content with that, since we strive for a higher reliability, aiming at least at two sources confirming the same ASN. Therefore, we first turn to `RISwhois`. If `RISwhois` gives us the very same ASN as `Traceroute`, we end the procedure and output that ASN as the correct one. If that's not the case we call the standard `whois`, which acts as the final referee. If it confirms one of the two ASNs previously obtained by `Traceroute` and `RISwhois`, then we obtain a 2/3 majority vote and declare that as the correct ASN. If, unfortunately, `whois` returns a third ASN, different from those obtained with `Traceroute` and `RISwhois`, the procedure is considered to fail. The whole procedure is reported as Algorithm 1.

At this point, we have the full list of ISPs administering the hops along the path to the end cloud server. We can then identify the hops belonging to the cloud provider through the algorithm described as Algorithm 2.

### E. Cloud availability estimation

Now, we have hopefully identified where the probing packets actually enter the cloud. We have all the data needed to measure the actual cloud availability.

If we indicate by  $N_{in}$  the number of probing packets entering the cloud, i.e. making it to the first cloud hop, and by  $N_{out}$  the number of echoes actually returned from the cloud end

<sup>4</sup><https://www.ripe.net/analyse/archived-projects/ris-tools-web-interfaces>

---

### Algorithm 1: Identification of ASNs and ISPs

---

**Input:** Cloud Provider, AS-traceroute to Cloud Provider, selected hop

**Output:** ASN and ISP of selected hop

$ASN \leftarrow \text{null};$

$ISP \leftarrow \text{null};$

**if** all reports of selected hop are empty **then**

    return null;

**else**

$current\_report \leftarrow$  select not empty report from hop;

**while**  $ISP$  is null **do**

$ASN_1, IP_1 \leftarrow$  AS-traceroute ( $current\_report$ );

**if**  $ASN_1$  is not null **then**

$ASN_2, ISP_1 \leftarrow$  RISwhois( $IP_1$ );

**if**  $ASN_2$  is equal to  $ASN_1$  **then**

$ASN \leftarrow ASN_1;$

$ISP \leftarrow ISP_1;$

**else**

$ASN_{2.1}, ISP_{1.1} \leftarrow$  whois( $IP_1$ );

**if**  $ASN_{2.1}$  is equal to  $ASN_1$  **then**

$ASN \leftarrow ASN_1;$

$ISP \leftarrow ISP_{1.1};$

**else**

**if**  $ASN_{2.1}$  is equal to  $ASN_2$  **then**

$ASN \leftarrow ASN_2;$

**if**  $ISP_1$  is equal to  $ISP_{1.1}$  **then**

$ISP \leftarrow ISP_1;$

**else**

                        Error:  $current\_report \leftarrow$  select

                            another not empty report if

                            there's else break;

**else**

$ASN_2, ISP_1 \leftarrow$  RISwhois( $IP_1$ );

$ASN_{2.1}, ISP_{1.1} \leftarrow$  whois( $IP_1$ );

**if**  $ASN_2$  is equal to  $ASN_{2.1}$  **then**

$ASN \leftarrow ASN_2;$

**if**  $ISP_1$  is equal to  $ISP_{1.1}$  **then**

$ISP \leftarrow ISP_1;$

**else**

            Error:  $current\_report \leftarrow$  select another not

                empty report if there's else break;

    return ASN, ISP;

---

server (i.e., the final hop in the sequence of hops obtained with `traceroute`), our measurement of the cloud availability is

$$A = \frac{N_{out}}{N_{in}}. \quad (3)$$

This approach allows not to factor in the losses due to the network on the way to the cloud, since they do not enter the  $N_{in}$  term. A remaining limitation of the approach is that echoes actually sent back by the cloud end server may get lost due to network problems on the return path.

**Algorithm 2: Identification of the first cloud hop**


---

**Input:** Cloud Provider, AS-traceroute to Cloud Provider  
**Output:** Position of the first Cloud hop in AS-traceroute report

$ISP \leftarrow \text{null}$ ;  
 $hopPosition \leftarrow 1$ ;  
 $cloudStart \leftarrow \text{null}$ ;  
 $currentEntryPosition \leftarrow 1$ ;  
 $Table \leftarrow \text{empty key-value table}$ ;

**while** *there's hop in hopPosition of AS-traceroute* **do**  
   $selectedHop \leftarrow \text{select hop in hopPosition}$ ;  
   $ISP \leftarrow \text{Identification of ASNs and ISPs (selectedHop)}$ ;  
   $newTableEntry \leftarrow \text{append entry (hopPosition, ISP)}$ ;  
   $hopPosition \leftarrow hopPosition + 1$ ;

**while** *there's table entry in currentEntryPosition* **do**  
   $currentISP \leftarrow \text{getValue (currentEntryPosition)}$ ;  
  **if** *currentISP is equal to Cloud Provider* **then**  
    **if** *cloudStart is equal to null* **then**  
       $cloudStart \leftarrow currentEntryPosition$  ;  
    **else**  
       $cloudStart \leftarrow \text{null}$ ;  
   $currentEntryPosition \leftarrow currentEntryPosition + 1$ ;

**return**  $cloudStart$ ;

---

## V. EXPERIMENTAL RESULTS

We have applied the procedure described in Section IV to three major cloud providers. In this section, we report the results.

Our aim is to assess two different things: the dependability of our procedure and the availability of cloud providers. The latter is of course meaningful if our procedure possesses the former feature.

For both purposes we considered three major cloud providers: Google Cloud, Amazon AWS, and Rackspace (all of them are included in the survey reported in [42]). We performed 50 tests for each of them, for a total of 150 tests. Each test consisted in sending probing packets over a period of 8 hours, going through the procedure described in Section IV, and assessing whether the cloud has responded (i.e., it is working) or not. The overall duration of test was therefore 400 hours for each provider.

The dependability issue is crucial. We have to be sure that the procedure works under real conditions and may be employed routinely. We stress the fact that our procedure requires neither the use of special software nor restricted information.

In Table I, we report the test results. Reporting an ASN as outcome means that we were able to get an ASN for all the hops along the path from source to destination, excluding from the count those hops that did not respond (for which we have of course no elements at all to infer their ASN). Overall, we get the ASN for the whole path roughly in 95% of cases; this result represents a good advance over the 90% declared in

Outcome	Frequency [%]
ASN (2/2 confidence level)	90.66
ASN (2/3 confidence level)	4.00
No ASN	5.34

TABLE I: Test results for procedure dependability assessment

the reference paper [40]. In the remaining 5% of cases there were some hops for which, though they did respond, we were not able to get a consensus over the ASN. For non-responding hops, a possible way to dispel the darkness is suggested again in [40]: if a non-responding hop is located along the path between two responding hops exhibiting the same ASN, then it is safe to assign that same ASN to the non-responding path. This solution would leave out just those non-responding hops located at the border between two ASes.

However, the final aim is to correctly assign outages, and we need to identify the first hop belonging to the cloud provider. In that case, non-responding hops may represent a problem, since they can actually be those belonging to the cloud provider. In our battery of tests, we were unable to identify the first cloud hop in 30% of cases, practically all due to Amazon (where the identification procedure failed in 45 out of 50 tests). Though this may appear as a disappointing performance, we must consider that a) it concerns a single provider; b) it is a matter of policy, which may be circumvented by arrangements between the cloud provider and the third-party organisation in charge of conducting the availability measurement (for example, if allowing for such measurements to be conducted on the basis of an agreement, Amazon could enable its routers to respond to probing packets sent by the authorised organisation, e.g. by recognising its IP addresses).

Once we have assessed that the procedure can be routinely carried out, we can employ it to assess the actual availability of the cloud. We consider the three major cloud providers that we have already mentioned: Google, Amazon, and Rackspace. We have carried out daily tests (lasting 8 hours) over 30 days, identifying the first hop belonging to the cloud provider and correctly assigning packet losses.

The results are shown in Table II, where we see that all three providers offer an availability better than 3 nines (actually quite close to 4 nines). There are two questions that naturally arise after these results:

- Is the contribution of the network relevant in availability assessment?
- Do these results confirm previous measurement campaigns?

The former question impacts the relevance of our measurement procedure. If the contribution of the network were negligible, there would be no interest in providing a measurement procedure capable of distinguishing between the failures taking place on the net and on the cloud. We see in Table II that in two out of three cases the network losses are at least twice as large as those due to the cloud. Even in the case of Rackspace they are all but negligible. In the absence of any loss attribution procedure the observed availability would

Measurement	Google Cloud	Amazon AWS	Rackspace
# probing packets	864 000	864 000	864 000
# packets reaching cloud	863 750	863 711	863 899
# lost packets (network)	250	289	101
# lost packets (cloud)	125	104	195
Availability	99.9855%	99.9879 %	99.9774%

TABLE II: Availability measurements

be 99.9566%, 99.9545%, and 99.9657% respectively. The difference between those values and those in Table II may look negligible, but we must not forget that we are talking about figures very close to 100% anyway and a difference as low as 0.01% is significant in this context (see an account of router availability issues in [43]). In the case of Google Cloud the actual difference appears to be 0.0289%, which would amount to 152 min (roughly 2 hours and a half) more downtime in a year, which is not negligible, given the quality-of-service expectations of customers. In addition, making a bundle of network and cloud losses would significantly alter the relative performances of the three cloud providers: Rackspace, ranking third in the correct measurement, would jump to the first place if we decided not to distinguish between the two sources of loss.

We can now turn to the latter question: how do these results compare with past measurement campaigns? We have reminded in the Introduction that there's not a host of measurement campaigns on cloud performance. However, a procedure like ours, which does not attribute to the cloud losses that are not its fault, naturally results in better performance figures for the cloud. Actually, though the results reported here are by no means exhaustive and conclusive, the availability look much better than was previously feared [18].

## VI. CONCLUSIONS

Our procedure allows us to assign the cloud the outages that are actually due to it and excluding those due to the network. It increases the accuracy of existing availability measurement procedures. The procedure can be conducted from any third-party vantage point and may be safely employed to assess the compliance of cloud providers with SLAs. The early results of its application show that the availability of cloud providers may be significantly underestimated.

Some limitations need to be addressed, though. A major limitation is that the non-response rate may be significant and must be reduced, since that prevents from obtaining a full view of the ISPs along the way. Though this can be achieved by way of agreements between measuring parties and cloud providers, it is too optimistic to hope for a 100% response rate. A second limitation is that network losses on the path back to the source may still cause the availability to be underestimated.

## REFERENCES

- [1] M. M. Qiu, Y. Zhou, and C. Wang, "Systematic analysis of public cloud service level agreements and related business values," in *Services Computing (SCC), 2013 IEEE International Conference on*. Santa Clara, CA, USA: IEEE, 2013, pp. 729–736.
- [2] S. A. Baset, "Cloud SLAs: present and future," *ACM SIGOPS Operating Systems Review*, vol. 46, no. 2, pp. 57–66, 2012.
- [3] M. Alhamad, T. Dillon, and E. Chang, "Conceptual sla framework for cloud computing," in *Digital Ecosystems and Technologies (DEST), 2010 4th IEEE International Conference on*. Dubai, United Arab Emirates: IEEE, 2010, pp. 606–610.
- [4] B. Varghese and R. Buyya, "Next generation cloud computing: New trends and research directions," *Future Generation Computer Systems*, vol. 79, pp. 849–861, 2018.
- [5] R. Buyya, S. N. Srirama, G. Casale, R. Calheiros, Y. Simmhan, B. Varghese, E. Gelenbe, B. Javadi, L. M. Vaquero, M. A. Netto *et al.*, "A manifesto for future generation cloud computing: Research directions for the next decade," *ACM Computing Surveys (CSUR)*, vol. 51, no. 5, p. 105, 2018.
- [6] M. Cingue, S. Russo, C. Esposito, K.-K. R. Choo, F. Free-Nelson, and C. A. Kamhoua, "Cloud reliability: Possible sources of security and legal issues?" *IEEE Cloud Computing*, vol. 5, no. 3, pp. 31–38, 2018.
- [7] L. Fiondella, S. S. Gokhale, and V. B. Mendiratta, "Cloud incident data: An empirical analysis," in *Cloud Engineering (IC2E), 2013 IEEE International Conference on*. San Francisco, California, USA: IEEE, 2013, pp. 241–249.
- [8] P. T. Endo, G. L. Santos, D. Rosendo, D. M. Gomes, A. Moreira, J. Kelner, D. Sadok, G. E. Gonçalves, and M. Mahloo, "Minimizing and managing cloud failures," *Computer*, vol. 50, no. 11, pp. 86–90, 2017.
- [9] R. Nachiappan, B. Javadi, R. N. Calheiros, and K. M. Matawie, "Cloud storage reliability for big data applications: A state of the art survey," *Journal of Network and Computer Applications*, vol. 97, pp. 35–47, 2017.
- [10] M. R. Mesbahi, A. M. Rahmani, and M. Hosseinzadeh, "Highly reliable architecture using the 80/20 rule in cloud computing datacenters," *Future Generation Computer Systems*, vol. 77, pp. 77–86, 2017.
- [11] B. Liu, X. Chang, Z. Han, K. Trivedi, and R. J. Rodríguez, "Model-based sensitivity analysis of iaas cloud availability," *Future Generation Computer Systems*, vol. 83, pp. 1–13, 2018.
- [12] H. Adamu, B. Mohammed, A. B. Maina, A. Cullen, H. Ugail, and I. Awan, "An approach to failure prediction in a cloud based environment," in *Future Internet of Things and Cloud (FiCloud), 2017 IEEE 5th International Conference on*. Prague, Czech Republic: IEEE, 2017, pp. 191–197.
- [13] Q. Lin, K. Hsieh, Y. Dang, H. Zhang, K. Sui, Y. Xu, J.-G. Lou, C. Li, Y. Wu, R. Yao *et al.*, "Predicting node failure in cloud service systems," in *Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*. Lake Buena Vista, Florida: ACM, 2018, pp. 480–490.
- [14] T. Labidi, A. Mtibaa, W. Gaaloul, S. Tata, and F. Gargouri, "Cloud sla modeling and monitoring," in *Services Computing (SCC), 2017 IEEE International Conference on*. Honolulu, HI, USA: IEEE, 2017, pp. 338–345.
- [15] F. Nawaz, O. K. Hussain, N. Janjua, and E. Chang, "A proactive event-driven approach for dynamic qos compliance in cloud of things," in *Proceedings of the International Conference on Web Intelligence*. Leipzig, Germany: ACM, 2017, pp. 971–975.
- [16] S. Alboghdady, S. Winter, A. Taha, H. Zhang, and N. Suri, "C'mon: Monitoring the compliance of cloud services to contracted properties," in *Proceedings of the 12th International Conference on Availability, Reliability and Security*. Reggio Calabria, Italy: ACM, 2017, p. 36.
- [17] H. J. Syed, A. Gani, R. W. Ahmad, M. K. Khan, and A. I. A. Ahmed, "Cloud monitoring: A review, taxonomy, and open research issues," *Journal of Network and Computer Applications*, 2017.
- [18] M. Naldi, "The availability of cloud-based services: Is it living up to its promise?" in *9th International Conference on the Design of Reliable Communication Networks, DRCN 2013, Budapest, Hungary*, 2013, pp. 282–289.
- [19] J. Dunne and D. Malone, "Obscured by the cloud: A resource allocation framework to model cloud outage events," *Journal of Systems and Software*, vol. 131, pp. 218–229, 2017.
- [20] M. Naldi, "Accuracy of third-party cloud availability estimation through ICMP," in *Telecommunications and Signal Processing (TSP), 2016 39th International Conference on*. Vienna, Austria: IEEE, 2016, pp. 40–43.
- [21] —, "ICMP-based third-party estimation of cloud availability," *International Journal of Advances in Telecommunications, Electrotechnics, Signals and Systems*, vol. 6, no. 1, pp. 11–18, 2017.

- [22] Z. Hu, L. Zhu, C. Ardi, E. Katz-Bassett, H. V. Madhyastha, J. Heidemann, and M. Yu, "The need for end-to-end evaluation of cloud availability," in *International Conference on Passive and Active Network Measurement*. Springer, 2014, pp. 119–130.
- [23] Y. Jadeja and K. Modi, "Cloud computing-concepts, architecture and challenges," in *Computing, Electronics and Electrical Technologies (ICCEET), 2012 International Conference on*. IEEE, 2012, pp. 877–880.
- [24] M. D. Dikaiakos, D. Katsaros, P. Mehra, G. Pallis, and A. Vakali, "Cloud computing: Distributed internet computing for it and scientific research," *IEEE Internet computing*, vol. 13, no. 5, 2009.
- [25] W. K. Hon and C. Millard, "Banking in the cloud: Part 3—contractual issues," *Computer Law & Security Review*, vol. 34, no. 3, pp. 595–614, 2018.
- [26] S. B. Rahi, S. Bisui, and S. C. Misra, "Identifying critical challenges in the adoption of cloud-based services," *International Journal of Communication Systems*, vol. 30, no. 12, p. e3261, 2017.
- [27] AA.VV., "Cost of data center outages," The Ponemon Institute, Tech. Rep., 2016.
- [28] G. Hogben and A. Pannetrat, "Mutant apples: a critical examination of cloud SLA availability definitions," in *Cloud Computing Technology and Science (CloudCom), 2013 IEEE 5th International Conference on*, vol. 1. Bristol, United Kingdom: IEEE, 2013, pp. 379–386.
- [29] K. R. Fall and W. R. Stevens, *TCP/IP illustrated, volume 1: The protocols*. addison-Wesley, 2011.
- [30] G. F. Lyon, *Nmap network scanning: The official Nmap project guide to network discovery and security scanning*. Insecure, 2009.
- [31] F. Viger, B. Augustin, X. Cuvellier, C. Magnien, M. Latapy, T. Friedman, and R. Teixeira, "Detection, understanding, and prevention of traceroute measurement artifacts," *Computer networks*, vol. 52, no. 5, pp. 998–1018, 2008.
- [32] B. Yao, R. Viswanathan, F. Chang, and D. Waddington, "Topology inference in the presence of anonymous routers," in *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies. IEEE Societies*, vol. 1. San Francisco California, USA: IEEE, 2003, pp. 353–363.
- [33] R. Govindan and H. Tangmunarunkit, "Heuristics for internet map discovery," in *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 3. Tel Aviv, Israel: IEEE, 2000, pp. 1371–1380.
- [34] J. Sommers, P. Barford, and B. Eriksson, "On the prevalence and characteristics of mpls deployments in the open internet," in *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*. Berlin, Germany: ACM, 2011, pp. 445–462.
- [35] B. Donnet, M. Luckie, P. Mérindol, and J.-J. Pansiot, "Revealing mpls tunnels obscured from traceroute," *ACM SIGCOMM Computer Communication Review*, vol. 42, no. 2, pp. 87–93, 2012.
- [36] S. Savage *et al.*, "Sting: A TCP-based Network Measurement Tool." in *USENIX Symposium on Internet Technologies and Systems*, vol. 2. Boulder, Colorado, USA, 1999, pp. 7–7.
- [37] M. Luckie, Y. Hyun, and B. Huffaker, "Traceroute probe method and forward ip path inference," in *Proceedings of the 8th ACM SIGCOMM conference on Internet measurement*. Vouliagmeni, Greece: ACM, 2008, pp. 311–324.
- [38] R. Motamedi, R. Rejaie, and W. Willinger, "A survey of techniques for internet topology discovery," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 1044–1065, 2015.
- [39] W. De Donato, P. Marchetta, and A. Pescapé, "A hands-on look at active probing using the ip prespecified timestamp option," in *International Conference on Passive and Active Network Measurement*. Vienna, Austria: Springer, 2012, pp. 189–199.
- [40] Z. M. Mao, J. Rexford, J. Wang, and R. H. Katz, "Towards an accurate as-level traceroute tool," in *Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*. Karlsruhe, Germany: ACM, 2003, pp. 365–378.
- [41] M. Luckie, A. Dhamdhare, B. Huffaker, D. Clark *et al.*, "bdrmap: inference of borders between ip networks," in *Proceedings of the 2016 Internet Measurement Conference*. Santa Monica, California, USA: ACM, 2016, pp. 381–396.
- [42] M. Naldi and L. Mastroeni, "Cloud storage pricing: A comparison of current practices," in *Proceedings of the 2013 International Workshop on Hot Topics in Cloud Services*, ser. HotTopiCS '13. New York, NY, USA: ACM, 2013, pp. 27–34.
- [43] A. Agapi, K. Birman, R. M. Broberg, C. Cotton, T. Kielmann, M. Millnert, R. Payne, R. Surton, and R. Van Renesse, "Routers for the cloud: Can the internet achieve 5-nines availability?" *IEEE Internet Computing*, vol. 15, no. 5, p. 72, 2011.

# Advances in Network Systems and Applications

**M**ODERN network systems encompass a wide range of solutions and technologies, including wireless and wired networks, network systems, services and applications. This results in numerous active research areas oriented towards various technical, scientific and social aspects of network systems and applications. The primary objective of track Network Systems and Applications conference track is to

group network-related technical sessions and promote synergy between different fields of network-related research.

The track currently consists of technical sessions:

- ANSA—Advances in Network Systems and Applications
- IoT-ECAW'19—3<sup>rd</sup> Workshop on Internet of Things—Enablers, Challenges and Applications



# Formalization of Software Risk Assessment Results in Legal Metrology Based on ISO/IEC 18045 Vulnerability Analysis

Marko Esche, Felix Salwiczek  
Physikalisch-Technische Bundesanstalt,  
Abbestraße 2-12, 10587 Berlin, Germany  
Email: {marko.esche, felix.salwiczek}@ptb.de

Federico Grasso Toro  
Federal Institute of Metrology METAS,  
Lindenweg 50, 3003 Bern-Wabern, Switzerland  
Email: federico.grasso@metas.ch

**Abstract**—The Measuring Instruments Directive sets down essential requirements for measuring instruments subject to legal control in the EU. It dictates that a risk assessment must be performed before such instruments are put on the market. Because of the increasing importance of software in measuring instruments, a specifically tailored software risk assessment method has been previously developed and published. Related research has been done on graphical representation of threats by attack probability trees. The final stage is to formalize the method to prove its reproducibility and resilience against the complexity of future instruments. To this end, an inter-institutional comparison of the method is currently being conducted across national metrology institutes, while the weighing equipment manufacturers’ association CECIP has provided a new measuring instrument concept, as a significant example of complex instruments. Based on the results of the comparison, a template to formalize the software risk assessment method is proposed here.

## I. INTRODUCTION

WHEN MEASURING results obtained from a measuring instrument are used to determine the price to pay for a certain good (such as water, heat, petrol, electricity) in the EU, said instrument is subject to the Measuring Instruments Directive (MID) 2014/32/EU [1]. The essential requirements of the MID include software requirements for protection against corruption, see L 96/173 in [1]. In addition, the MID defines conformity assessment procedures which an instrument has to pass before being made available on the common market. In the frame of most of these assessment procedures, manufacturers are required to conduct a risk assessment demonstrating that their product fulfils the essential requirements. To aid manufacturers with this task, PTB (Germany’s national metrology institute, one notified conformity assessment body for the MID) has developed a software risk assessment procedure [2]. This procedure, specifically tailored to the needs of legal metrology, i.e. the economic sector of measurements subject to legal control, is employed by PTB when performing conformity assessments. To harmonize conformity assessment practice across Europe, the European Cooperation

This work was conducted within the frame of work package 4 of the European Metrology Cloud Project.

in Legal Metrology (WELMEC) Working Group 7 “Software” is currently investigating this software risk assessment method in the frame of an inter-institutional comparison. The aim is to demonstrate the objectiveness of the procedure and the reproducibility of its results. If needed, it is intended to amend the procedure to achieve both goals. To ensure impartial results, generic abstract instruments are used for the comparison. Initial findings indicate that producing objective assessment results for today’s simple instruments should be feasible. However, future complex systems will pose a bigger challenge. Most importantly, the simple representation of the assessment result in the form of a single risk score simplifies the assessment process too much. Therefore, it is proposed to improve the investigated method by formalizing the recording of its results, by means of a risk assessment template. Since the procedure closely follows the vulnerability analysis of ISO/IEC 18045 [3], the outcome of this paper should be useful to all assessment procedures (such as ETSI TS 102 165-1 [4]) that are based on the same standard. The remainder of the paper is structured as follows. The basic principles of the risk assessment procedure are recapitulated in Section II. Section III details the inter-institutional comparison, the examined generic measuring instruments and describes challenges derived from the results of the comparison. The proposed solution by means of a formalized risk assessment template is detailed in Section IV. Section V summarizes the paper.

## II. BASIC PRINCIPLES OF THE RISK ASSESSMENT PROCEDURE

As mentioned in the introduction, manufacturers of measuring instruments shall perform and document a risk assessment of their instruments before submitting a prototype to a NB for conformity assessment, in accordance with Module B (type evaluation) of the MID. To aid manufacturers and NBs in this task, a procedure was developed and published in [2]. With the aim of providing an objective procedure to generate reproducible results, the method is based on the international standards ISO/IEC 27005 [5] and ISO/IEC 18045 [3]. ISO/IEC 27005 provides a principle description of the risk assessment process consisting of three phases:

TABLE I  
TOE RESISTANCE OF MEASURING INSTRUMENTS TO ATTACKS AND  
ASSOCIATED PROBABILITY SCORE [3].

Sum of points	TOE resistance	Probability score
0-9	No rating	5
10-13	Basic	4
14-19	Enhanced Basic	3
20-24	Moderate	2
$\geq 24$	High	1

### A. Risk Identification

During risk identification, unwanted events (so-called threats to assets) are defined based on "legal and regulatory requirements, and contractual obligations". Such assets can be derived from the essential requirements given in Annex I of the MID. For convenience reasons, only two such assets are examined here. One asset is the measurement result with the associated security property authenticity, since the MID prohibits the use of measurement results that do not originate from a certified measuring instrument. The other asset is the software critical for the measurement purpose, which shall not be modified or replaced. Therefore, such software can be assigned the security properties integrity and authenticity. A list of all assets applicable to legal metrology is given in [2].

### B. Risk Estimation

During risk estimation, threats are assigned a quantitative or qualitative risk measure. One possibility to calculate such a measure is given by ISO/IEC 27005 itself, where "risk is a combination of the consequences that would follow from the occurrence of an unwanted event and the likelihood of the occurrence of the event." If unwanted events (threats), have been defined properly, they can be assigned an impact score between 0 (no effect) and 1 (all measurement results affected), signifying the severity of the consequences. The method from [2] uses a score of  $\frac{1}{3}$  if only one result is affected by the threat. In addition, a measure for the probability of occurrence is needed. This can be estimated by evaluation of different actions (attack vectors) that an attacker needs to implement for the threat to be realized. The vulnerability analysis provided in Part 2 of ISO/IEC 18045 [3] constitutes one possibility to quantify the probability of occurrence for such attack vectors by means of point scores assigned in the following categories:

- Elapsed Time (0-19 points)
- Expertise (0-8 points)
- Knowledge of the Target of Evaluation (0-11 points)
- Window of Opportunity (0-10 points)
- Equipment (0-9 points)

An example for a fully evaluated attack vector with assigned scores is given in Table III. The calculated sum score can be mapped to a target of evaluation (TOE) resistance, see [3], and an equivalent probability score between 1 and 5, see Table I. The third column is not part of the original table presented in [3]. Afterwards, by multiplying impact and probability score a risk score can be obtained which will be in the range between 1 (very low risk) and 5 (very high risk).

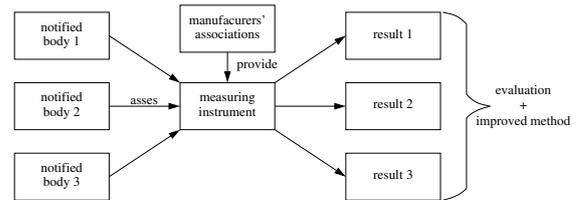


Fig. 1. Anticipated workflow of the risk assessment inter-institutional comparison and its projected outcome.

### C. Risk Evaluation

During risk evaluation, the calculated risk is put into the context of the field of application for the assessed instrument. Estimated risks are prioritized and a cut off point for the risk assessment is defined. In addition, an initial list of risks to be mitigated by order of importance is produced. As a rule of thumb, PTB will ask manufacturers who have obtained a risk score of 4 or 5 for their instrument to implement additional protective measures. Once these have been implemented, the phases of risk estimation and risk evaluation (including amendments, where necessary) are repeated until the risk score is reduced to 3 or lower. Since attack vectors for real-world measuring instruments might become very complex, they can be decomposed by means of Attack Probability Trees (AtPT), see [6]. These AtPTs can be used by an assessor to subdivide any given attack vector, evaluate the sub-goals and to find the attack probability score for the original complex attack vector.

## III. INTER-INSTITUTIONAL COMPARISON AND IDENTIFICATION OF CHALLENGES

Since conformity assessment bodies all across Europe are faced with the challenge of interpreting and evaluating the results of risk assessments, WELMEC Working Group 7 has decided to examine the procedure developed by PTB more closely by means of an inter-institutional comparison with five different NBs. To start the comparison, assessors from these NBs took part in a training exercise. The training covered both the basic procedure [2] as well as AtPTs [6]. Afterwards, see Subsection III-A, two generic measuring instruments were selected for all partners to assess. Subsection III-B describes the examined threats and initial findings. Figure 1 illustrates the workflow of the inter-institutional comparison.

### A. Description of Generic Reference Instruments

The first instrument assessed is a complex cloud-based measuring system proposed by CECIP, the European weighing instruments manufacturers' association. WELMEC Working Group 7 anticipates that such systems will be the norm in legal metrology in the near future. The system consists of a number of sensors subject to legal control that send data to a processing software running in the instrument manufacturer's own cloud, see Figure 2. The cloud offers data storage and a display server (DSP). The DSP sends measurement results to different display devices, e.g. smart phones or general-purpose printers. Communication between the components is

realized via Wi-Fi with WPA encryption. Additionally, all transferred data are protected by CRC-16 codes to ensure integrity of transmitted data. Three kinds of users are foreseen for the cloud: administrators with full privileges, maintenance personal with access to log files and backend users. In case data are lost during transmission, all sending devices have sufficiently large buffers for retransmission. Two categories of display devices are established, namely "full control" and "receive only", where the prior devices are only accessible to a trustworthy user group. A full system description will be published by CECIP in a future paper. Due to the complexity of the cloud-based instrument and the resulting increased probability for assessment errors, it was agreed to also use a second simpler generic instrument.

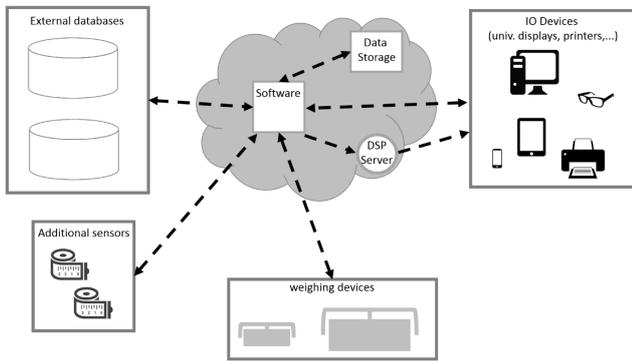


Fig. 2. General structure of the generic complex cloud-based system provided by CECIP, the European weighing instruments manufacturers' association.

The weighbridge depicted in Figure 3 is an automatic weighing instrument designed for weighing cargo transported on a truck. The measurement is started through the terminal's GUI consisting of an LCD and eight buttons. The measurement result is directly shown on the LCD. Two load cells measure the weight of front and rear axle of the truck. Two evaluator units interpret the output of the load cells. These units then communicate with the terminal where the final measurement result is computed. Evaluator units and terminal are based on microprocessors, data can be read from the terminal via RS485 or exported to a USB stick. The terminal checks the authenticity of all other units at startup by requesting a CRC-16 of their firmware based on a secret start vector. Legally relevant parameters and software are stored in the terminal unit on a hardware-protected flash memory. All software on the system is subject to legal control. All connections within the system are physically sealed.

*B. Experimental Results of the Inter-institutional comparison*

To narrow down the scope of the comparison, it was agreed to examine only two threats for both instruments, although in

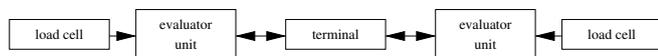


Fig. 3. Components of the generic automatic weighing instrument.

TABLE II  
ASSESSMENT RESULTS PROVIDED BY DIFFERENT NBS.

Threat	NB	Impact	Probability score	Risk
complex cloud-based system				
T1	NB1	1	4	4
	NB2	1	1	1
	NB3	1	4	4
	NB4	1	4	4
	NB5	1	3	3
T2	NB1	1	2	2
	NB2	1	1	1
	NB3	1	5	5
	NB5	1	3	3
simple measuring instrument				
T1	NB1	1/3	2	1
	NB3	1	3	3
	NB4	1	3	3
	NB5	1	3	3
	T2	NB1	1	1
NB3		1	3	3
NB4		1	3	3
NB5		1	3	3

principle, all assets derived from the MID would need to be taken into account:

- T1: An attacker introduces false measurement results into the measuring instrument.
- T2: An attacker modifies or replaces the software critical for the measuring task.

All NBs were asked to identify at least one attack vector per threat per measuring instrument and to evaluate that attack vector as described in Section II. The outcome is a list of point scores for the five mentioned categories. The sum score results in a probability score, see Table I, which produces a risk score when multiplied with the identified impact. Table II summarizes the results provided the NBs for threats T1 and T2. Not all NBs assessed both threats for both instruments. The results from different NBs for threat T1 for the cloud-based system are visualized in Figure 4. For this threat, all NBs selected an attack vector with a permanent effect (impact score of 1). Despite varying sum scores (11 to 17), the probability and risk scores for NB1, NB3, NB4 and NB5 are very close to each other due to the range of sum scores allowed by Table I. The only exception is the attack vector selected by NB2, which appears to be more difficult to implement than the others. When comparing results from NB4 and NB5, another property of the ISO/IEC 18045 vulnerability analysis becomes apparent: A larger score for expertise might be compensated by a smaller score for time, since a layman may take longer to implement a certain attack than an expert. While the results for T1 might suggest that consistent results can be easily obtained by different assessors, the results for threat T2 prove otherwise, see Figure 5. All four NBs concluded that the chosen attack vector would have a permanent effect. For all other scores, the results vary widely. Consequentially, probability and risk scores also differ. One reason for this variability is the imprecisely formulated threat T2, which allows either a partial modification or a complete replacement

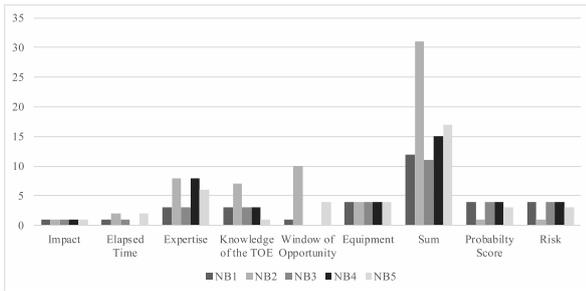


Fig. 4. Results for the complex cloud-based measuring system, evaluation of threat T1 (introduction of false measurement results).

of the software. While a small modification will require less expertise and time (attack examined by NB3), a full redevelopment will require more expertise, time etc. (attack examined by NB1). It is concluded, that properly documented attack vectors and precisely defined threats are key to comparing risk assessment results obtained by different parties. Moreover, assessment results for the cloud-based system do not depend on the perspective of the examiner alone, but also on the chosen attack vector. If an attacker aims to introduce false measurement results by providing them with a valid CRC-16 from a trustworthy source, this will be much less difficult than manipulating data within the WPA-protected Wi-Fi. To solve this disambiguity, all NBs were also asked to perform software risk assessments for the much simpler weighbridge instrument, detailed in Subsection III-A. The results obtained for threat T1 are depicted in Figure 6. NB3, NB4 and NB5 chose to examine attack vectors with a permanent effect, e.g. replacement of a sensor. Again, the point scores vary depending on the selected attacker profile (layman with restricted knowledge vs. expert with publicly available knowledge). Nevertheless, all three NBs arrive at similar sum and probability scores, resulting in identical risk scores. NB1 has examined an alternative attack vector requiring repetition for each measurement (reduced impact of  $\frac{1}{3}$ ). Since this attack also appears to be more complex (writing a specialized software vs. installing a sensor) the risk score obtained is much lower. In this regard, a set of evaluated reference attack vectors could reduce the assessor's required effort and harmonize the outcome of different assessments

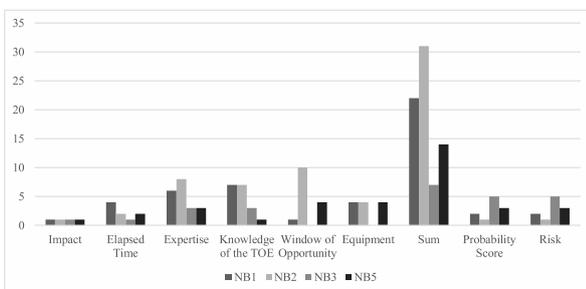


Fig. 5. Results for the complex cloud-based measuring system, evaluation of threat T2 (modification or replacement of software critical for the measurement).

performed for the same instrument. Concerning threat T2, the same effect can be observed, see Figure 7. While NB3 and NB5 focused on the simple modification of existing software, NB1 expected attackers to implement new software and then replace the original. The differences are due to the different focus of the attack vectors and an imprecisely formulated threat. This could be avoided by formulating individual threats per identified asset and requiring assessors to document chosen attack vectors and their effects for later comparison.

### C. Main Challenges

The justification - for rejecting certain assessed attack vectors as unlikely or for quantifying certain scores as wrong - was provided by a review session among the NBs involved. In practice, discussions between NBs about risk analyses provided by different manufacturers are unlikely to happen. Moreover, new examiners may not be familiar with these findings and will be facing the same challenges. As shown in Subsection III-B, an objective comparison of risk assessments is only possible if certain prerequisites are fulfilled:

- Instructions for new evaluators on how to assess risks according to the standard shall be readily available.
- Examples for evaluation of common attack vectors to reduce the workload for evaluators shall be supplied.
- Proper documentation of the complete attack vector and justification for the evaluation shall be required of all assessors for better comparability of assessment results.

Section IV addresses all three by providing a formalized framework, by means of a software risk assessment template.

### IV. FORMALIZATION OF RISK ASSESSMENT RESULTS

Instructions on how to perform a vulnerability analysis are provided by Part 2 clause B.4.2.2 ff of ISO/IEC 18045. Since the method discussed here is based on that standard, the same instructions may be used when performing software risk assessments in legal metrology. However, the standard's guidance is intended for all fields of IT security and is thus kept very general. In the template proposed, a shorter method description is included focused on the needs of legal metrology. Thereby, it is ensured that all assessors are aware of all steps to be performed. The workflow of the template is shown in Figure 8. The template includes a list of all assets

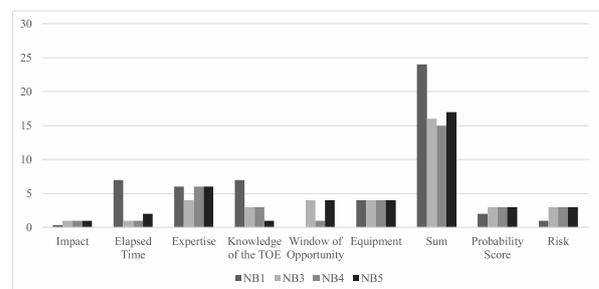


Fig. 6. Results for the simple weighbridge, evaluation of threat T1 (introduction of false measurement results).

TABLE III  
EXAMPLE FOR A FULLY EVALUATED ATTACK VECTOR.

Attack vector	Time	Expe- ri- se	Knowl- edge	Window of opport.	Equip- ment	Justification
Attacker constructs fake results from datasets protected by a CRC32 with a secret start vector.	0	3	3	0	0	Assumed attacker: customer. CRC is a linear operation on binary vectors, an XOR-connection of two datasets automatically produces a third dataset with correct CRC. This can be calculated with standard software by a proficient user. No window of opportunity needed. The CRC is described in the manual.

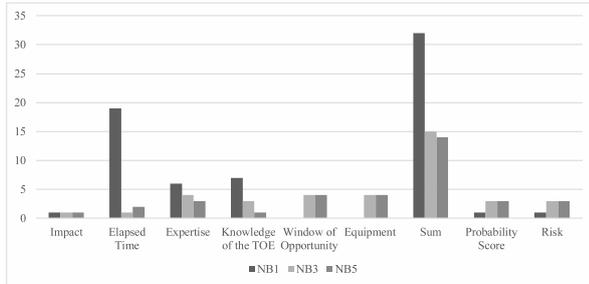


Fig. 7. Results for the simple weighbridge, evaluation of threat T2 (modification or replacement of software critical for the measurement).

and their security properties derived from the MID. These are accompanied by explanations of the assets and examples on how an attacker might invalidate their security properties. The assessor is required to select the applicable assets from the list and to formulate the relevant threats precisely. Even though the identification of attack vectors cannot be standardized, the template provides assessors with some assistance. A number of evaluated reference attack vectors are given, see Table III for an example. Also, assessors are offered the possibility to use AtPTs to decompose attack vectors to facilitate the evaluation. Most importantly, the template requires all assessors to provide justification for each point score alongside the evaluated attack vector. If sufficient details and justification are provided, discussion about assessment results will no longer be necessary, unless a different assessor takes issue with the point score assigned to a specific attribute. In this case, an AtPT can be used to decompose the

attack until no room for argument is left. The template will not guarantee uniform results, but if the guidance remarks are observed, there will be sufficient documentation to successfully argue in favor or against a certain assessment result. The template can be found under the following link: [https://www.ptb.de/cms/fileadmin/internet/fachabteilungen/abteilung\\_8/8.5\\_metrologische\\_informationstechnik/8.51/Risk\\_Assessment\\_Template\\_v11.docx](https://www.ptb.de/cms/fileadmin/internet/fachabteilungen/abteilung_8/8.5_metrologische_informationstechnik/8.51/Risk_Assessment_Template_v11.docx)

V. SUMMARY

As long as software risk assessment depends on human creativity and judgement, the resulting risk scores will always be biased. Nevertheless, detailed guidance on the assessment steps together with proper documentation of all steps of the assessment may serve as a basis to make software risk assessment results more easily comparable. The vulnerability analysis of ISO/IEC 18045 already provides general remarks on the workflow and on the point scores for specific attributes of assessed attack vectors. These were mapped to the needs of the legal metrology community and augmented by specific detailed examples to help assessors with repetitive tasks. The experimental findings from the inter-institutional comparison and the suggested risk assessment template, should be applicable to any group planning to implement ISO/IEC 18045 vulnerability analysis. To validate the template, WELMEC Working Group 7 is currently performing a second stage of risk assessments using the new template.

REFERENCES

- [1] “Directive 2014/32/EU of the European Parliament and of the Council of 26 February 2014 on the harmonisation of the laws of the Member States relating to the making available on the market of measuring instruments,” European Union, Council of the European Union ; European Parliament, Directive, February 2014.
- [2] M. Esche and F. Thiel, “Software risk assessment for measuring instruments in legal metrology,” in *Proceedings of the Federated Conference on Computer Science and Information Systems*, Lodz, Poland, September 2015. doi: <http://dx.doi.org/10.15439/978-83-60810-66-8 pp. 1113–1123>.
- [3] “ISO/IEC 18045:2008 Common Methodology for Information Technology Security Evaluation,” International Organization for Standardization, Geneva, CH, Standard, September 2008, Version 3.1 Revision 4.
- [4] “ETSI TS 102 165-1 Telecommunications and Internet converged Services and Protocols for Advanced Networking; Methods and protocols; Part 1: Method and proforma for Threat, Risk, Vulnerability Analysis,” European Telecommunications Standards Institute, Sophia Antipolis Cedex, FR, Standard, March 2011, v4.2.3.
- [5] “ISO/IEC 27005:2011(e) Information technology - Security techniques - Information security risk management,” International Organization for Standardization, Geneva, CH, Standard, June 2011.
- [6] M. Esche, F. Grasso Toro, and F. Thiel, “Representation of attacker motivation in software risk assessment using attack probability trees,” in *Proceedings of the Federated Conference on Computer Science and Information Systems*, Prague, Czech Republic, September 2017. doi: <http://dx.doi.org/10.15439/2017F112 pp. 763–771>.

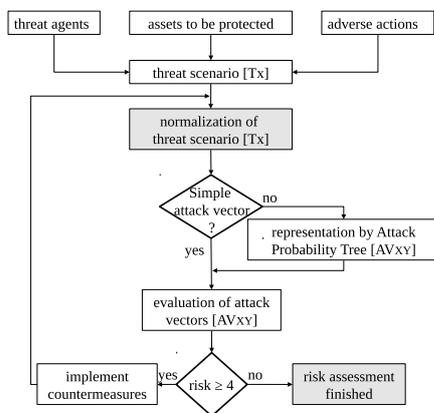


Fig. 8. Template workflow mirrors risk assessment procedure [2].



# FANE: A Firewall Appliance for the Smart Home

Christoph Haar

Hochschule für Telekommunikation  
in Leipzig

Gustav-Freytag-Straße 43-45, 04277 Leipzig, Germany  
Email: haar@hft-leipzig.de

Erik Buchmann

Hochschule für Telekommunikation  
in Leipzig

Gustav-Freytag-Straße 43-45, 04277 Leipzig, Germany  
Email: buchmann@hft-leipzig.de

**Abstract**—With the advent of the Internet of Things (IoT), many domestic devices have been equipped with information technology. By connecting IoT devices with each other and with the Internet, Smart Home installations exist that allow the automation of complex household tasks. A popular example is Google Nest that controls cooling, heating and home security. However, Smart Home users are tempted to neglect that such IoT devices pose IT-Security risks. Examples like the Mirai malware have already shown that insecure IoT devices can be used for large-scale network attacks. Thus, it is important to adapt security approaches to Smart Home installations. In this paper, we introduce FANE, our concept for a Firewall Appliance for Smart Home installations. FANE makes a few realistic assumptions on the network segmentation and the communication profile of IoT devices. This allows FANE to learn firewall rules automatically. Our prototypical implementation indicates that FANE can secure a wide range of IoT devices without requiring network-security expertise from the Smart Home user.

## I. INTRODUCTION

IN THE last years, the proliferation of Smart Home installations has gained momentum. Today, the consumer market offers a huge number of different Internet-of-Things (IoT) devices.

Smart thermostats, cameras, speakers and even toothbrushes contain information technology that connects the IoT device over the Internet with cloud services or other IoT devices. For example, IoT devices from the Google Nest family [1] provide a straightforward, user-friendly way to control heating, cooling and home security. Smart speakers like Amazon Alexa [2] allow to control many daily activities via voice control. From the perspective of the manufacturers, the Smart Home concept allows new business models, e.g., to sell new product features as digital upgrades for IoT devices.

On the other hand, consumers might be tempted to overlook that the IoT devices pose an IT-Security risk. For example, the lifetime of a traditional security camera ends when the device is broken. In contrast, the lifetime of an IoT security camera that connects over the Internet should come to an end when its manufacturer discontinues security updates, even if the IoT security camera is still working. Otherwise, the IoT security camera might end up as, say, part of the Mirai bot network, which consisted of approx. 500,000 devices in 2016 [3].

From the perspective of a consumer without in-depth expertise of network security, it is next to impossible to find out if the IoT devices present in a Smart Home installation are subject to attacks over the Internet. In this paper, we

explore options to integrate a firewall into typical Smart Home installation that can detect and deter such attacks. This is challenging, since the firewall must be compliant with the typical modes of use of a Smart Home installation, and a consumer cannot be expected to evaluate firewall rules or identify false alarms. On the other hand, the IoT devices used differ from general-purpose devices such as smartphones and desktop computers. This might allow for pre-configuration to some extent.

In particular, we make the following contributions:

- 1) We systematically compare the lifecycle of a classical firewall with the lifecycle of IoT devices in a typical Smart Home installation.
- 2) We propose FANE, a Firewall Appliance on a Wi-Fi bridge in Smart Home installations.
- 3) We describe a proof-of-concept implementation of FANE based on a Raspberry Pi, and we evaluate it with three different IoT devices.

We show that it is possible to develop a generic IT-Security concept for IoT devices in a Smart Home installation by making few realistic assumptions, e.g., the IoT network segment is only used by single-purpose IoT devices, which do not fundamentally change their communication profiles. We have implemented this security concept in FANE. Our evaluation indicates that FANE can secure the IoT network segment without requiring the user to possess network-security expertise.

**Paper structure:** In Section II, we review related work. In Section III we provide a problem statement. We describe FANE in Section IV, followed by a proof-of-concept implementation in Section V and an experimental evaluation in Section VI. Section VII concludes.

## II. RELATED WORK

In this section we provide a brief definition of Internet of Things and Smart Home, and we introduce related work on firewalls, firewall management and approaches to generate firewall rules automatically.

### A. Internet of Things and Smart Homes

The "Internet of Things" (IoT) refers to physical appliances, which have been equipped with information technology in order to connect them with other devices directly or over the Internet [4]. IoT includes a wide range of appliances, from

connected cars over smart buildings to connected machinery in an Industry 4.0 setting. The concept "Smart Home" narrows down this range to devices that let end users to control, monitor or access everyday objects of the daily routine [5].

### B. Security Challenges

To assess the security properties of Smart Home installations, it is important consider the basic security challenges that occur in installations of IoT devices. One study [6] lists six major security issues:

**Identity and Authentication:** In IoT environments, numerous devices need to authenticate each other in order to provide trustable services. Thus, reliable techniques for identification and authentication are needed.

**Access Control:** To create new services it is necessary to aggregate data from different providers. This is challenging, because in typical IoT scenarios each provider has its own access control policy.

**Protocol and Network Security:** If IoT devices communicate with each other in a distributed network architecture, distributed schemes for key management are needed.

**Privacy:** The Smart Home concept means that numerous IoT devices monitor the actions of its users in order to devise meaningful responses. Thus, privacy very important from a user perspective.

**Trust and Governance:** In IoT architectures there are two dimensions of trust. The first dimension is between users and their IoT devices. The other dimension is between the IoT devices. Device A needs to trust the accuracy and integrity of the data produced by device B. Data governance goes in the same direction, in a sense of data and access governance.

**Fault Tolerance:** Mechanisms for fault tolerance need to be established to counteract faulty or tampered devices.

Other studies [7] list similar challenges.

### C. Firewalls State of the Art

Firewalls are able to control and log the network traffic based on rules set by an administrator or security expert. In literature different firewall generations are distinguished [8]. 1st generation firewalls are known as packet filters which operates on the transport layer. The filtering is based on source and destination IP addresses, ports and protocols. 2nd generation firewalls are also operating on the transport layer and they are known as stateful packet inspection. State tables are used to keep track of the network traffic and filtering is based on state and context of packets. 3rd generation firewalls are operating on the application level and require different proxies for each service. The proxy acts as a middleman between source and destination to reestablish a new session. Current firewall technologies are called next generation firewalls. These next generation firewalls are looking deep into packets and combine traditional firewall technologies with network filtering capabilities on the application level [9]. However, all these generations have in common that an expert is needed to define rules or check them for correctness which motivates our new approach.

### D. Firewall Lifecycle

Traditionally, a firewall must be part of the IT-Security process, as described by the ISO 270xx standards family [10], the German BSI Grundschutz Standard 200-2 [11] or the ITIL process for security Management [12]. The IT-Security process starts with a IT-Security policy that has been passed by the management. Based on this policy, business objectives, the assets to be protected and a risk classification can be identified. Subsequently, measures can be defined and implemented that restrict the IT-Security risks to acceptable levels. In the following, the effectiveness of these measures needs to be monitored. Based on this information, corrective actions can be planned and executed [13]. Note that all process steps require a person with IT-Security expertise, which cooperates with various IT experts from the operations department.

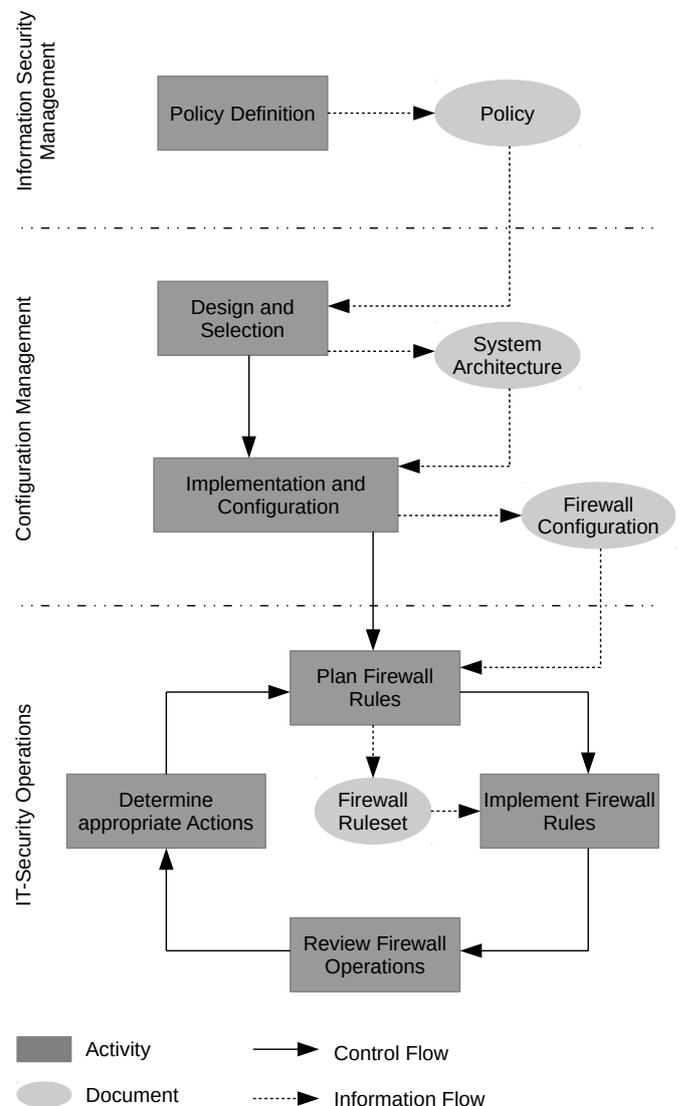


Fig. 1. Traditional Firewall-Lifecycle

A firewall fits into the IT-Security process [14] as shown in Figure 1. In the Information-Security Management phase, the management defines a security policy based on company-wide security objectives. This policy is independent from technical realities. Based on the security policy, an IT-Security expert designs the architecture of the firewall system and selects the firewall system components. In the implementation and configuration phase, the IT-Security expert adapts the firewall system to the system architecture with its network segments, hosts and applications. This includes a preliminary set of firewall rules that define which network packets are allowed to pass the firewall. In the next step, a plan-do-check-act cycle takes place where the firewall rules are designed, implemented, reviewed and improved in a repetitive way. Typically, this cycle is part of IT-Security operations. It allows to adjust the firewall rules to changes such as new business applications, hosts moving from one network segment to another one, or in case of detecting new attacks. Note that not only the management of firewall rules is a cyclic process, but also the IT-Security process. If the management observes that the security policy is ineffective, this policy can be changed as well, and it has an impact on all design decisions further down the IT-Security process chain.

#### E. Firewall Rules

It is a labor-intensive task for a domain expert to create a rule set for firewalls manually. One option to obtain firewall rules (semi-)automatically is to use data mining or machine learning on a training set consisting of network packets. This option is based on the assumption all user applications operate as intended while the training set is recorded. Respective approaches [15]–[17] have been proposed for Intrusion Detection systems, but might be adaptable to firewalls as well. By using k-Means, C4.5 decision tree algorithms, Naive Bayes classifier, Neural Networks or Support Vector Machines, it is possible to derive common characteristics of allowed network connections. Those characteristics could be translated into firewall rules. It is also possible to generate firewall rules by mining the firewall log [18] instead of a dump of network packets. However, all approaches require an IT-Security expert to decide which generated rules are relevant to meet the security requirements, and the quality of the generated rules still needs further research.

A different option to generate firewall rules is to deduce them from a formal specification of security requirements by using argumentation logic [19]. This approach allows to automatically obtain a detailed, comprehensive set of rules from a high-level specification. However, creating a specification of the security requirements for a certain system architecture still requires expert knowledge in IT-Security.

### III. PROBLEM STATEMENT

In this section, we explore the differences between traditional firewalls and firewalls needed for IoT devices in a Smart Home installation. In addition, we derive requirements for a Smart Home firewall.

#### A. Does a Firewall Fit into the Smart Home Concept?

To find out in which ways traditional firewall use cases differ from Smart Home use cases, we consider the modes of use, network architecture, application scenario, user roles and information technology used.

*a) Modes of Use:* A firewall is an access control mechanism that allows or blocks network traffic between two network segments that have different security properties [20], e.g., an internal network and the Internet that is open for anybody. The firewall enforces a set of firewall rules that allow or prohibit network packets to travel from one segment into the other one. The firewall rules depend on the use cases that are executed over both network segments. For example, a business workflow "Answer customer requests" might require that a set of machines in the internal network is allowed to send and receive email to/from the Internet. Thus, firewall rules must be defined by a network-security expert with domain knowledge. If the workflows, the applications or the segment boundaries are changed, the expert must adapt the firewall rules as well. Traditionally, firewalls are tailored for complex multi-purpose scenarios where the hosts execute numerous different applications that change over time.

Smart Home use cases are fundamentally different [21]: A typical IoT device is a physical object that has been extended with information technology to improve its usefulness. For example, a smart toothbrush [22] can tell its user if a tooth has gone unbrushed. Thus, IoT devices are constructed for a single purpose that does not change over time. It only makes sense to install a toothbrush control software on a smart toothbrush. As a result, IoT devices are single-purpose objects. If the device is not needed any more, it will be disposed.

*b) Network Architecture:* Firewalls depend on the network segmentation. With traditional use cases, a network installation might contain multiple segments protected by multiple firewalls. A prominent example is a perimeter network [20], which contains assets such as Web servers that must be accessible from an external network. Two sets of firewall rules protect the perimeter network against the external network and the internal network against the perimeter and the external network. However, the number and architecture of the network segments might be individually different for each network installation.

In contrast, a typical Smart Home installation with IoT devices produces three network segments with different security properties: (a) the untrusted Internet, (b) the home network with trusted devices such as the user's laptop and printer, and (c) an IoT network segment that contains all IoT devices. Since the IoT device and its software comes as an integrated package, the user has little options to influence the security of the IoT device, e.g., by disabling unused network protocols or by removing unused software functionality. Thus, the IoT network segment should be separated from the home network [23], which is used for sensible tasks such as online banking or online shopping. All devices in the IoT network segment can be expected to require an Internet connection, to

provide a service, to obtain updates and upgrades, to allow a remote control via smartphone app, etc.

c) *Application Scenario*: Firewalls follow the IT-Security lifecycle, as explained in Section II. Based on a general security policy that has been defined from a management perspective, a network-security expert defines the position of the firewall(s) in the network architecture and a set of firewall rules. By using a plan-do-check-act-cycle, the firewall rules as well as the firewall hard- and software must be constantly monitored, evaluated and adapted to changes in the IT infrastructure.

On the opposite side, one of the fundamental principles of the Smart Home concept is to let IoT devices use sensors to observe its environment, in order learn appropriate actions with a minimum of user interaction and without requiring the user to scrutinize the operations of the IoT device on a regular basis. For example, the nest thermostat observes the temperature preferences of its user and if he or she is at home, and controls the heating system accordingly. Furthermore, the duration of use of IoT devices is an one-dimensional process that starts with the deployment of a device and ends with its disposal, just like non-smart devices [24], i.e., it does not follow a periodic lifecycle where it is constantly monitored and improved. For example, a smart light switch never changes its function, and it cannot be adapted to different needs.

d) *User Roles*: Setting up a traditional firewall typically requires three distinct roles: The role "Information Security Management" defines a security policy by considering the assets and (business) objectives that are relevant for a certain part of the IT infrastructure. Based on the policy, the role "Configuration Management" designs a firewall system, selects appropriate firewall components, and provides an initial installation and configuration of the system. Finally, a role "IT-Security Operation" constantly monitors and improves the firewall system, both on the level of the firewall rules and of the firewall hard- and software.

In contrast, an IoT device for a Smart Home usually is pre-configured by the manufacturer for typical use cases. The end user can deploy and configure the IoT device with minimal efforts, does not need to monitor it later on and does not need expert knowledge.

e) *Information Technology*: IoT devices make use of network protocols which have been well established. They use Linux-based operating systems, Cloud resources and Open Source programming libraries. The network security of IoT devices is based on mechanisms for encryption, certification and signatures that have been in use for years. Thus, from a technical point of view, off-the-shelf firewalls can be directly used to control the network traffic of IoT devices.

### B. Problem Definition

From a technical perspective, it would be a simple exercise for a network security expert to set up a firewall that controls the network traffic of an IoT device. However, this procedure conflicts with the general understanding how IoT devices should operate in a Smart Home. Thus, a firewall for Smart

Homes must differ in the following properties from traditional firewalls:

- P1** The firewall must be usable without expert knowledge.
- P2** The firewall must fit to the durations of use of Smart Home components.
- P3** The firewall must operate in a way that is typical for IoT devices in the Smart Home.

**P1** implies not only that the configuration and installation of a firewall in a Smart Home must not require network security expertise. It also means that a user cannot be expected to tell false alarms from real alarms, or to decide if a certain firewall rule is applicable to the home network. From **P2** it follows that such a firewall must deal with IoT devices that are bought once for a certain purpose and never change its basic properties until disposal, and it must operate in the same way. Furthermore, the firewall must operate in the same way. **P3** means that a firewall in a Smart Home needs to operate without permanent care from the user, i.e., it must monitor the network traffic, deduce meaningful firewall rules and provide appropriate reactions to forbidden network packets.

We have ruled out a cloud-based approach [25], [26] that externalizes the firewall to a trusted third party on the Internet. Although such an approach might fulfil the properties described, it requires a permanent Internet connection. In addition, a cloud-based firewall would transfer security-relevant information into the cloud. Thus, both the Internet connection of the firewall and the trusted third party would be a valuable target for an attacker.

## IV. FANE: A FIREWALL APPLIANCE

In this section, we introduce FANE, a concept for a Firewall Appliance that is compatible with the Smart Home paradigm.

### A. Network Architecture

A firewall separates network segments with different security properties. Typical IoT devices do not allow its user to observe security properties, and to configure security-related aspects, such as disabling unused functions. Furthermore, an IoT device is designed to be used like a classical, non-smart device, i.e., its users are tempted to forget that the device might pose IT-Security risks. For this reason, IoT devices should be placed in network segments that are isolated from all other network segments of the Smart Home.

Thus, FANE operates as a Wi-Fi bridge that connects the IoT network segment to the Internet and includes a firewall, as shown in Figure 2. The IoT network segment only contains single-purpose IoT devices, and the Wi-Fi bridge is the only connection of the IoT segment to other network segments and the Internet. We observe that this allows us to specify the security concept in advance.

### B. Security Concept

From Section III it follows that a traditional firewall approach is complex, because the underlying network segmentation and the processes executed over the boundaries of these segments are complex, too, and might change from

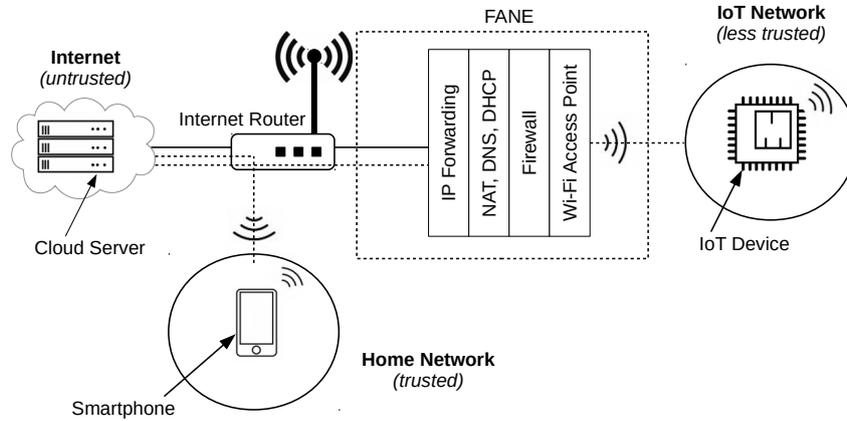


Fig. 2. System Architecture

time to time if a new software is installed on a device in the network. With our network architecture, we have reduced this complexity. We only need to consider three kinds of communication activities:

- An IoT device wants to communicate with a server on the Internet. For example, a smart thermostat wants to communicate with the user's smartphone, which is mediated over a cloud service.
- An IoT device wants to communicate with a device in another network segment. For example, the user installs a control application on a laptop to configure the smart thermostat.
- An IoT device wants to communicate with another IoT device in the same network segment. For example, our smart thermostat wants to directly communicate with the smart air condition.

Since FANE operates as a bridge to the Internet, only the first two kinds of communication have to be monitored, and the security properties of the endpoints of the communication can be specified at production-time of FANE: The open Internet is insecure by default, the IoT devices are less secure, and the devices in other network segments of the Smart Home are trustworthy. This allows to pre-configure the security concept of FANE in advance, i.e., it does not need a user with network-security expertise (Property **P1**):

- 1) No device on the Internet is allowed to open a network connection to the IoT network segment.
- 2) An IoT device is allowed to open a connection to the Internet, if this is part of its normal operation.

- 3) An IoT device is allowed to open a connection to devices in other (trusted) network segments of the Smart Home, if this is part of its normal operation.
- 4) A device from a trusted segment is allowed to open connections to the IoT network segment.
- 5) IoT devices are allowed to open connections to other devices in the IoT network segment.

### C. Smart Home Firewall Operations

FANE has to meet conflicting requirements: It must meet the expectations provided by Smart Home components (**P2**). In particular, this means that FANE must operate without constant supervision (**P3**). At the same time, as a security component it must not neglect the IT-Security process, including a plan-do-check-act cycle to refine firewall rules. However, this must be possible without requiring the user to possess expert knowledge (**P1**).

We circumvent these conflicts, as shown in Figure 3): We distinguish between pre-configuration management and Smart Home operations. Because we restrict FANE to the network architecture described in Subsection IV-A, the policy definition, the firewall design and a baseline configuration of firewall rules can be done at pre-configuration time. Thus, we shift the initial parts of the IT-Security process into the responsibility of the Smart Home firewall manufacturer who possess IT-Security expertise. Furthermore, we propose to automate the configuration and the plan-do-check-act cycle in a way that it's phases can be started without expert knowledge at operation time. Finally, we define a process step in a way that the user is informed when an IT-Security expert is needed.

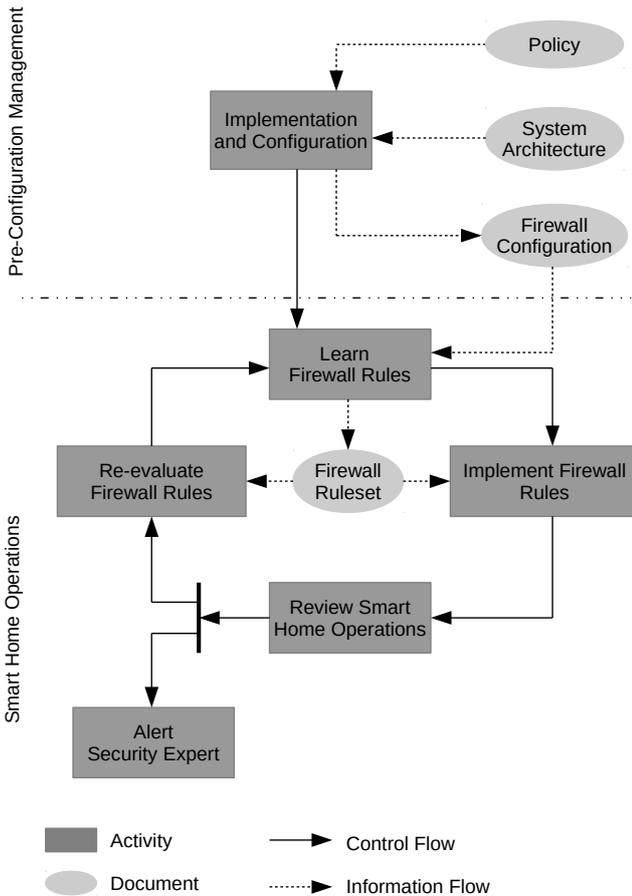


Fig. 3. FANE Operations

#### D. User Interaction

After having defined the operations of FANE, we can define the user interactions needed. Observe that no interaction requires expert knowledge (Property **P1**). FANE comes as an IoT device that runs out-of-the-box after being connected to a power outlet and the Internet.

When FANE is connected to the Smart Home installation for the first time or if new IoT devices are added, the user can tell FANE to **learn** new firewall rules by observing the network packets of the IoT devices. Assume an IoT device uses a functionality that has not been used during the learning stage, or the device has been updated and a new network connection is now blocked by FANE. In this case, the user has the option to let FANE **re-evaluate** the rule set. That is, FANE executes a learning stage on a certain device with the option to discard rules that have been learned before. The rules from the security concept (Subsection IV-B) cannot be discarded.

If FANE blocks a large number of network packets per time-interval, it generates an **alert**. The alert shows the user that immediate action needs to be taken, i.e., something happens

that cannot be handled automatically by FANE. For example, the IoT network segment might face a denial-of-service attack from the Internet, or an IoT device has been taken over and tries to connect to the attacker's command and control server on the Internet. In such cases, the user might decide to call the customer support of the IoT device, or ask an IT-Security expert for further investigations.

#### V. PROOF-OF-CONCEPT IMPLEMENTATION

In this section, we describe the software and hardware components of our FANE prototype, how FANE learns firewall rules and in which way it interacts with the user.

##### A. Our FANE prototype

We have realized FANE on the basis of a Raspberry Pi, which executes several linux shell scripts to configure and operate an iptables packet filter (see Subsection II-C). Figure 4 illustrates our hardware configuration and the main software packages used.

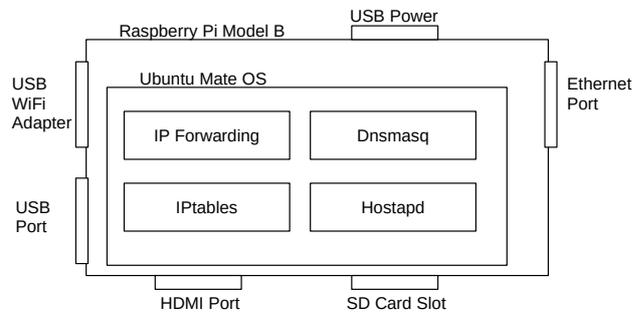


Fig. 4. Our FANE prototype

*a) Hardware:* From Section IV it follows that FANE must provide a Wi-Fi access point that creates a network segment for IoT devices. The IoT devices might want to communicate with other devices in the same segment, the home network segment and the Internet. Thus, FANE must be connected to the Internet, and its firewall must control all incoming and outgoing network packets of the IoT network segment.

We have implemented this approach on a third-generation Raspberry Pi model B. This is a credit-card sized single board computer containing a quad-core processor with 1.2GHz, 1 GB main memory and various network and connection interfaces. Because the on-board Wi-Fi chip cannot be configured as a Wi-Fi access point, we have connected an external Wi-Fi module via USB. We have used a 32 GB SD Card for permanent storage.

We would need only two switches to initiate the learning- and re-evaluation stage of the user interface, and one LED indicating an alert. The IT-Security expert, which might be needed to handle serious attacks on the IoT network segment, would be able to obtain firewall logs and other information by using an SSH connection. This way, our FANE prototype

costs less than 60 EUR. However, to ease development we have used an external USB keyboard and a LCD monitor.

b) *Software*: We have used the Ubuntu Mate Linux operating system as a basis of our software configuration. On top of a minimal OS installation, we need the following software packages and services:

- awk (script language to edit text files)
- cron (timed execution of processes)
- dnsmasq (DHCP client and DNS cache)
- hostapd (Wi-Fi access point)
- inotify-tools (monitor changes in files)
- iptables (network address translation and firewall)
- tcpdump (record network packets)

By configuring the Ethernet interface *eth0* as a DHCP client, our Raspberry Pi can be connected to any Internet router without further configuration. We have configured the *wlan0* interface with a static IP address and subnet mask, and we have configured it as a Wi-Fi access point by using *hostapd*. Our Smart Home firewall must act as a bridge between *eth0* (Internet) and *wlan0* (Wi-Fi segment for IoT devices). Thus, we have used *iptables* and *sysctl* to activate IP forwarding, including network-address translation and masquerading. With *dnsmasq*, we have realized a DHCP service.

### B. Learning Firewall Rules

For our FANE prototype, we have used a straightforward approach to learn firewall rules. For more elaborate approaches, see Section II. The learning stage consists of two phases, a *monitoring phase* and a *rule generation phase*. We assume that all network traffic recorded during the monitoring phase is allowed, i.e., we assume that no IoT device has been manipulated or attacked before the monitoring phase ends.

When FANE is connected to power and Internet for the first time, or if the user wants FANE to learn new rules, it enters the monitoring phase for a certain period of time. In this phase, FANE waits for new IoT devices connecting to the access point, and logs the network packets. We have implemented this phase as follows:

At boot time, a *cron* task with the time prefix *@reboot* starts a script that finds out if the set of firewall rules is the one that has been pre-configured from the security concept (Subsection IV-B). Alternatively, a user command starts the monitoring phase manually. In the monitoring phase, FANE uses the monitoring tool *inotify* to find out if the *dhcp leases* file changes. This indicates new devices using the access point. In this case, *inotify* executes a script that obtains the IP address of the device from *dhcp leases*. At the same time, FANE uses *tcpdump* to create a log file containing all network packets sent or received during the monitoring phase.

At the end of the monitoring phase, FANE stops *tcpdump* and enters the rule generation phase. In this phase, FANE parses the log file from *tcpdump* into firewall rules according to the IP addresses of the IoT devices that have used the access point in the monitoring phase. In particular, FANE uses a *sed* command to filter the log for incoming and outgoing IP addresses and ports. This set of addresses and ports is

reduced to unique entries in a second step. The odd lines in Figure 5 show, how the set of addresses and ports looks like after FANE has removed surplus information and duplicates from the log file. In a third step, a shell scripts parses the remaining addresses and ports into firewall rules that allow such packets for the *iptables* chain "FORWARD". The odd lines in Figure 5 illustrate this step. We have used the *iptables* policy "DROP", i.e., FANE drops all packets that are not allowed by the rules generated.

```

1 15:23:18 IP 10.200.65.101.1080 > 35.158.162.95.80:
2 iptables -A FORWARD -s 10.200.65.101 -sport
   1024:65535 -d 35.158.162.95 -dport 80
   -p tcp -j ACCEPT
3 15:23:22 IP 10.200.65.101.8553 > 35.157.158.75.1883:
4 iptables -A FORWARD -s 10.200.65.101 -sport
   1024:65535 -d 35.157.158.75 -dport 1024:65535
   -p tcp -j ACCEPT
5 15:24:36 IP 10.200.65.101.8653 > 35.156.40.103.1883:
6 iptables -A FORWARD -s 10.200.65.101 -sport
   1024:65535 -d 35.156.40.103 -dport 1024:65535
   -p tcp -j ACCEPT
7 15:25:07 IP 10.200.65.101.8554 > 35.157.255.122.80:
8 iptables -A FORWARD -s 10.200.65.101 -sport
   1024:65535 -d 35.157.255.122 -dport 80
   -p tcp -j ACCEPT

```

Fig. 5. Firewall rules learned from an adjusted packet log

Note that this procedure can be extended easily to extended firewall features, e.g., to include the *iptables* options for stateful inspection. At the end of the rule generation phase, FANE installs the rules and is ready for operation.

If an IoT device is not working properly, if a new IoT device is added to the IoT network segment or if an existing device is used in a way it has never been used before, the user can order FANE to re-evaluate the rule set. In this case, the user has the option to discard rules from preceding learning procedures, and to re-start the monitoring- and rule-generation phase.

## VI. EXPERIMENTAL EVALUATION

In this section, we explore the applicability of FANE with three different Smart Home appliances.

### A. Setup

Figure 6 shows our experimental setup. FANE is directly connected to the Internet router, and its integrated access point spans a Wi-Fi network segment for IoT devices. The Internet router creates a Wi-Fi home network that connects a smartphone to the Internet. Different cloud services connect the smartphone to the IoT devices. A cloud service might use a load balancer, i.e., the IP addresses the IoT devices connect to might change from time to time.

We have tested three different devices, which communicate differently with a control app on the user's smartphone:

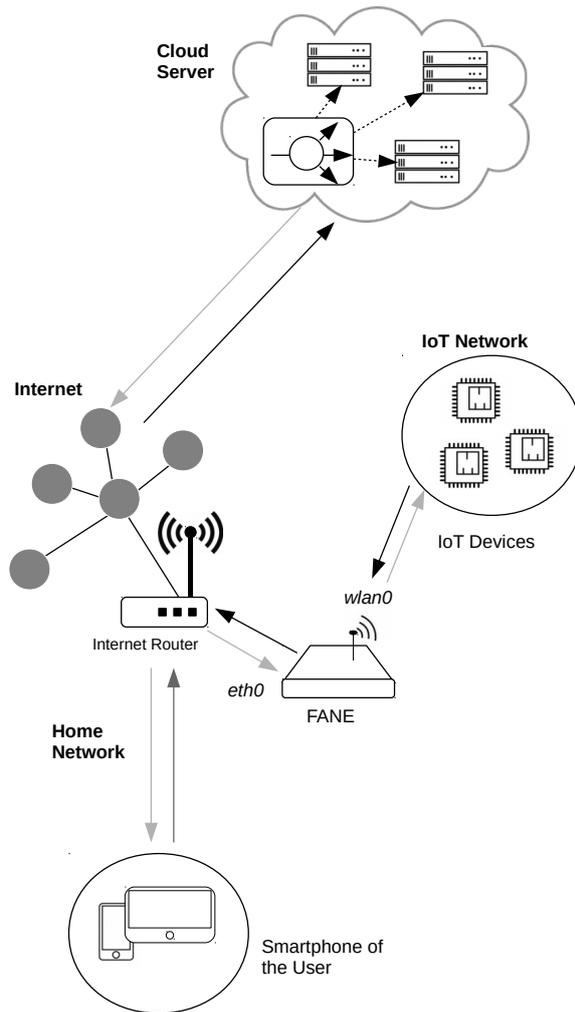


Fig. 6. Our experimental setup

- 1) An electrical IoT relay.
- 2) An IoT power outlet.
- 3) An IoT security camera.

The IoT devices do not communicate directly with each other, but with the user's smartphone and the Internet. Thus, for our experiments we do not need to preconfigure rule 5 from our security concept (see Subsection IV-B). We have configured each device for FANE's IoT network segment. We have used a monitoring phase of 40 minutes, and we have operated each device periodically during this phase. In the following, we briefly introduce each IoT device, and we describe what we have learned by using FANE as described.

### B. IoT Relay

Our first use case is an electrical relay "10A Wi-Fi smart switch", sold for less than 9 EUR, manufactured by Sonoff [27]. The IoT relay can be turned on or off via

smartphone app, which allows a technician to integrate non-smart electrical devices into a straightforward Smart Home installation. Sending commands from the app to the relay requires an Internet connection, i.e., there is no option to directly connect the smartphone app to the IoT device. After the relay is connected to the access point provided by FANE, and the user has installed the smartphone app, the relay is ready to use.

In our monitoring phase of 40 minutes, we have switched the relay on and off frequently via smartphone app for 10 minutes. After that, we have waited for a period of 20 minutes. Finally, we have operated the relay for further 10 minutes.

After completing the monitoring phase, FANE has written 1,800 lines in the packet log. All packets followed the TCP protocol and were sent/received to/from one singular IP address located at a dedicated server leased from Amazon. Thus, the rule generation phase has generated only one rule for in- and outgoing packets. The IoT relay was working properly after FANE has activated the firewall rule set generated. Figure 7 shows an example from the traffic log FANE has recorded from the IoT relay.

```

1 13:41:31.551813 IP 10.200.65.109.55147 >
   52.71.154.91.443: Flags [F.], ...
2 13:41:31.551870 IP 10.200.65.109.55145 >
   52.71.154.91.443: Flags [F.], ...
3 13:41:31.551914 IP 10.200.65.109.55161 >
   52.71.154.91.443: Flags [.], ...
4 13:41:31.668878 IP 52.71.154.91.443 >
   10.200.65.109.55161: Flags [.], ...
5 13:41:31.669239 IP 52.71.154.91.443 >
   10.200.65.109.55161: Flags [P.], ...

```

Fig. 7. Fragment of the packet log of the IoT relay

### C. IoT Power Outlet

Our second use case is an IoT power outlet "Smart Wi-Fi Socket Model SWA1", sold for 18 EUR, produced by Shenzhen Ligan Intelligent Technology [28]. Similarly to the IoT relay, the IoT power outlet can be turned on or off via smartphone app. In addition, it can be controlled with Amazon Alexa or Google Home, which allows to integrate non-smart electrical devices into an elaborate Smart Home concept without requiring a technician. Any command to the IoT power outlet is handled by a cloud service over the Internet.

In our monitoring phase, we have used the IoT power outlet via smartphone in the same way as the relay for 40 minutes. At the end of the monitoring phase, FANE has collected a packet log of approx. 2,600 lines, all of them TCP packets. The rule generation phase has generated rules that allow five

different IP addresses, all of them in the address range of the Amazon AWS cloud.

The IoT power outlet was fully operational after FANE has started to filter network connections. We have observed that only one of the five addresses in the firewall rule set was actually used to operate the outlet via smartphone app. We assume that some network connections are used only for analyzing customer behavior or similar purposes, i.e., blocking them would not reduce the functionality of the device.

#### D. IoT Security Camera

The most complex IoT device tested was a "720P HD IP Wireless security camera", sold for 37 EUR and manufactured by XinweiYa [29]. The IoT security camera sends a live video stream to the smartphone of the user. Furthermore, the smartphone app allows to restart the IoT security camera, and to rotate it around two axes. After connecting the IoT security camera to a power outlet, it can be configured with a smartphone app to use FANE's access point.

During our monitoring phase of 40 minutes, we have restarted the IoT security camera, we have let the IoT security camera sent a live video stream of 10 minutes to the smartphone, we have waited for 20 minutes, and we have restarted it again for another live stream of 10 minutes. After 40 minutes, FANE has collected 8 MB packet log of approx. 27,000 lines, most of them UDP packets.

The rule generation phase produces a rule set of 20 rules for this device. Those rules allow services like Network Time Protocol (NTP) or Domain Name System (DNS) as well as cloud services hosted on Amazon AWS, the Microsoft cloud and the Alibaba cloud.

We have observed that the IoT security camera was not working properly, after FANE started to filter network packets. Our investigations have shown that this due to a specific load balancer. The IP address of the load balancer was allowed by the firewall rule set generated. But the load balancer referred the IoT security camera frequently to IP addresses unknown to FANE. However, it would be possible to adapt the learning approach to cope with such a load balancer. For example, FANE could detect and accept IP addresses that are close by addresses that are already allowed by the rule set.

The packet log has also shown that the IoT security camera first tries to reach the smartphone app in the same network segment directly, via multicast. Thus, even if the IoT security camera makes use of the Internet connection, it might be able to provide its basic functionality without the Internet. From this observation we conclude that there might be options for FANE to distinguish between communication needed for the normal operation of an IoT device, and other communication needed for advertising purposes or usage analytics that can be blocked without undesired side-effects.

Finally, we have observed that the IoT security camera produces more network load by an order of magnitude than the other IoT devices tested. While this has slowed down the rule generation phase, it did not overstrain the IP forwarding capacity of our Raspberry Pi during normal operation.

#### E. Discussion

Our three use cases have provided evidence that a straightforward learning approach is applicable to many IoT devices used in Smart Home scenarios. Two of our three IoT devices remained fully operative after FANE has monitored the network activities of our devices for 40 minutes, and has subsequently generated and activated firewall rules. Furthermore, our observations have shown that it would be easily possible to extend our learning approach to consider load balancers. As there is no communication standard for IoT devices, it is problematic to generalize our findings to all IoT devices used in the Smart Home. However, using a cloud service seems to be typical for many use cases. Only network packets can pass FANE that are allowed by a specific rule. Thus, FANE increases the security of the Smart Home installation.

FANE operates without requiring the user to possess expert knowledge, by making three assumptions: First, the network segment created by FANE's access point contains IoT devices only. This allows to specify a security policy in advance, before FANE is delivered to the user. Second, the IoT devices operate as single-purpose appliances that do not fundamentally change their communication profiles. Due to this assumption, FANE can learn a rule set that remains stable over a long period of time, which makes it compatible with the Smart Home concept. Third, we assume that the IoT devices are working properly during the monitoring phase. This allows FANE to learn firewall rules unattended.

## VII. CONCLUSION

The last years have brought a plethora of Internet-of-Things (IoT) devices dedicated to Smart Home installations. While such IoT devices have numerous practical use cases, observations have shown that many of them come with IT-Security risks. For example, the Mirai botnet consisted of approx. 500,000 baby-phones, security cameras and other insecure IoT devices that were able to execute distributed denial-of-service attacks with 1 Tbit/s network bandwidth. However, typical Smart Home users do not possess the network-security knowledge needed to identify and deter attacks on IoT devices. Furthermore, the Smart Home concept encourages the users to leave IoT devices unattended for long periods of time.

In this paper, we have introduced FANE, our concept for a Firewall Appliance for Smart Home installations. FANE makes a few realistic assumptions on the network segmentation and the communication profile of IoT devices. This allows to pre-configure FANE with a generic security concept. It also enables FANE to learn firewall rules automatically by observing the network traffic of IoT devices.

Experiments with a prototypical implementation have provided evidence that FANE can secure ordinary IoT devices without requiring network-security expertise from the Smart Home user. Only one device was not working properly after FANE has activated its firewall rules due to a specific load balancer. However, this problem could be solved by accepting IP addresses close to addresses that FANE already knows.

## ACKNOWLEDGMENT

We would like to thank Eric Ilgunas for his exceptional work on realizing and evaluating the FANE prototype.

## REFERENCES

- [1] Nest Labs, *Nest*, <https://nest.com/>, Accessed: 2019-02-25.
- [2] Wareable Ltd., *Amazon Echo voice control*, <https://www.the-ambient.com/guides/best-amazon-alexa-commands-280>, Accessed: 2019-02-25.
- [3] C. Koliass, G. Kambourakis, A. Stavrou, and J. Voas, "Ddos in the iot: Mirai and other botnets," *Computer*, vol. 50, no. 7, pp. 80–84, 2017.
- [4] P. P. Gaikwad, J. P. Gabhane, and S. S. Golait, "A survey based on smart homes system using internet-of-things," in *2015 International Conference on Computation of Power, Energy, Information and Communication*, IEEE, 2015, pp. 0330–0335.
- [5] L. Jiang, D.-Y. Liu, and B. Yang, "Smart home research," in *Proceedings of 2004 International Conference on Machine Learning and Cybernetics (IEEE Cat. No. 04EX826)*, IEEE, vol. 2, 2004, pp. 659–663.
- [6] R. Roman, J. Zhou, and J. Lopez, "On the features and challenges of security and privacy in distributed internet of things," *Computer Networks*, vol. 57, no. 10, pp. 2266–2279, 2013.
- [7] T. Heer, O. Garcia-Morchon, R. Hummen, S. L. Keoh, S. S. Kumar, and K. Wehrle, "Security challenges in the ip-based internet of things," *Wireless Personal Communications*, vol. 61, no. 3, pp. 527–542, 2011.
- [8] K. Neupane, R. Haddad, and L. Chen, "Next generation firewall for network security: A survey," in *Southeast-Con 2018*, IEEE, 2018, pp. 1–6.
- [9] J. Surana, K. Singh, N. Bairagi, N. Mehto, and N. Jaiswal, "Survey on next generation firewall," *International Journal of Engineering Research and Development*, vol. 5, no. 2, pp. 984–988, 2017.
- [10] G. Disterer, "Iso/iec 27000, 27001 and 27002 for information security management," 2013.
- [11] Bundesamt für Sicherheit in der Informationstechnik, "BSI-Standard 200-2, IT-Grundschutz-Methodik," <https://www.bsi.bund.de>, 2017.
- [12] O. of Government Commerce, *Introduction to ITIL, The key to managing IT services*. Van Haren Publishing, 2005.
- [13] S. Fenz, G. Goluch, A. Ekelhart, B. Riedl, and E. Weippl, "Information security fortification by ontological mapping of the iso/iec 27001 standard," in *13th Pacific Rim International Symposium on Dependable Computing*, IEEE, 2007, pp. 381–388.
- [14] S. W. Lodin and C. L. Schuba, "Firewalls fend off invasions from the net," *IEEE spectrum*, vol. 35, no. 2, pp. 26–34, 1998.
- [15] K. Jaswal, P. Kumar, and S. Rawat, "Design and development of a prototype application for intrusion detection using data mining," in *2015 4th international conference on reliability, infocom technologies and optimization*, IEEE, 2015, pp. 1–6.
- [16] L. S. Parihar and A. Tiwari, "Survey on intrusion detection using data mining methods," *International Journal for Science and Advanced Research in Technology*, vol. 3, no. 12, pp. 342–7, 2016.
- [17] A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1153–1176, 2016.
- [18] K. Golnabi, R. K. Min, L. Khan, and E. Al-Shaer, "Analysis of firewall policy rules using data mining techniques," in *2006 IEEE/IFIP Network Operations and Management Symposium NOMS 2006*, IEEE, 2006, pp. 305–315.
- [19] A. K. Bandara, A. C. Kakas, E. C. Lupu, and A. Russo, "Using argumentation logic for firewall configuration management," in *2009 IFIP/IEEE International Symposium on Integrated Network Management*, IEEE, 2009, pp. 180–187.
- [20] D. B. Chapman, E. D. Zwicky, and D. Russell, *Building internet firewalls*. O'Reilly & Associates, Inc., 1995.
- [21] G. Kortuem, F. Kawsar, V. Sundramoorthy, D. Fitton, et al., "Smart objects as building blocks for the internet of things," *IEEE Internet Computing*, vol. 14, no. 1, pp. 44–51, 2009.
- [22] Procter & Gamble, *Oral-b genius electric toothbrushes*, <https://www.oralb.co.uk/en-gb/products/electric-toothbrushes/oral-b-genius>, Accessed: 2019-04-25.
- [23] N. Gupta, V. Naik, and S. Sengupta, "A firewall for internet of things," in *2017 9th International Conference on Communication Systems and Networks*, IEEE, 2017, pp. 411–412.
- [24] J. Stark, "Product lifecycle management," in *Product lifecycle management*, Springer, 2015.
- [25] A. R. Khakpour and A. X. Liu, "First step toward cloud-based firewalling," in *2012 IEEE 31st Symposium on Reliable Distributed Systems*, IEEE, 2012, pp. 41–50.
- [26] C. Modi, D. Patel, B. Borisaniya, H. Patel, A. Patel, and M. Rajarajan, "A survey of intrusion detection techniques in cloud," *Journal of network and computer applications*, vol. 36, no. 1, pp. 42–57, 2013.
- [27] ewelink, *Sonoff relay*, <http://ewelink.coolkit.cc>, Accessed: 2019-04-25.
- [28] lingansmart, *Power outlet*, <http://www.lingansmart.com>, Accessed: 2019-04-25.
- [29] XinweiYa Co.,Ltd., *Security camera*, <http://www.cctvgood.com>, Accessed: 2019-04-25.

# Standardized container virtualization approach for collecting host intrusion detection data

Martin Max Röhling, Martin Grimmer\*,  
Dennis Kreußel<sup>†</sup>, Jörn Hoffmann\*  
Leipzig University  
Ritterstraße 9-13, 04109 Leipzig  
Email: roehling@wifa.uni-leipzig.de

\*{grimmer, jhoffmann}@informatik.uni-leipzig.de  
<sup>†</sup>dnk0@protonmail.com

Bogdan Franczyk  
Leipzig University  
Grimmaische Straße 12, 04109 Leipzig  
Uniwersytet Ekonomiczny we Wrocławiu  
ul. Komandorska 118/120, 53-345 Wrocław  
Email: franczyk@wifa.uni-leipzig.de

**Abstract**—Anomaly-based Intrusion Detection Systems (IDS) can be instrumental in detecting attacks on IT systems. For evaluation and training of IDS, data sets containing samples of common security-scenarios are essential. Existing data sets are not sufficient for training modern IDS. This work introduces a new methodology for recording data that is useful in the context of intrusion detection. The approach presented is comprised of a system architecture as well as a novel framework for simulating security-related scenarios.

## I. INTRODUCTION

The current threat situation of the IT landscape makes it necessary to monitor systems and detect attacks at an early stage. Host-based Intrusion Detection Systems (HIDS) are important tools to inspect system calls and to analyze processes which are accessing systems. Especially anomaly-based HIDS are able to detect previously unknown attacks. These are trained in advance with normal behavior and detect deviant behavior in the event of an attack. The quality of an anomaly-based HIDS in relation to the detection and error rate is significantly linked to the quality of the training of these systems. In recent years, various data sets have been published to evaluate a HIDS. As it turns out all them have at least one serious problem [1]. In addition for comparability and evaluation, their metrics must be applied to a set of coherent standardized data sets. Thus, all existing data sets are not sufficiently applicable to design anomaly-based HIDS for the modern IT landscape. Especially when modern operating systems, multithreaded applications and concurrent communications are considered.

The methodology presented in this paper enables the simulation and comprehensive recording of normal and attack behavior with an high degree of detail. Further, we suggest plausible practices for implementing this approach. This includes a procedure to generate new data sets on current operating systems. The latter are suitable to develop and evaluate algorithms for today's state of the art anomaly-based HIDS.

This work was partly funded by the German Federal Ministry of Education and Research within the project “Explicit Privacy-Preserving Host Intrusion Detection System” (EXPLOIDS) (BMBF federation code 16KIS0522K) and “Competence Center for Scalable Data Services and Solutions” (ScADS) Dresden/Leipzig (BMBF federation code 01IS14014B).

## A. Relevance in practice and research

Our methodology can be used to record normal behavior from productive systems. This generates models based only on the data actually captured from live containers. This is the opposite to today's HIDS that often rely entirely on data captured from a staged lab environment. This way, a user can simulate a security violation by implementing a custom policy. This results in a transparent process, in which an IDS can be build and evaluated with an productive system in mind.

The procedure model serves research primarily with experimenting and evaluation of new IDS algorithms. In today's research, new algorithms are evaluated with outdated and incomplete data sets. Current data sets with extensive context enable new research and better evaluation of the algorithms.

## II. BACKGROUND AND RELATED WORK

Since 1998 data sets for training and comparison of HIDS have been published. The best known are: the DARPA Intrusion Detection Evaluation Data Set (KDD), from 1998 to 2000 [2] the data set of the University of New Mexico (UNM) from 1999, [3], [4], the data sets of the Australian Defence Force Academy, the ADFA-LD from 2013 [5], [6] and the NGIDS-DS from 2017 [7]. They share at least one of the problems described by Grimmer et al. [8] as shown in table I. These data sets consist of sequences of system calls. The ADFA-LD for example is relatively up-to-date, but it does not provide thread information, parameters and return values. In addition it contains system calls of a complete system with all its processes. In particular, it is therefore not possible to learn the normal behavior of a single program from it. Short examples for all four data sets can be seen in listing 1.

Unfortunately, the authors of the data sets omit many details about their implementation. There is little information about the general conditions under which the recordings have been made. This refers to information on the recording process, the tools and software versions used or the attack vectors engineered into the system. Moreover, the normal behavior and its origin are either not described at all or only insufficiently described. Pendleton and Xu describe in [1] an architecture

based on a syscall collector for data generation. For instrumentation the Intel tool "Pin" is used<sup>1</sup>. It allows applications to be extended with their own source code at runtime and to profile the application. This allows thread-based system call sequences and context information to be captured. The authors show this with the software example Firefox. Abed et al. describe in [9] a real-time IDS for passive monitoring of Linux containers using the tool strace. The evaluation takes place with the example of the database application MySQL in the normal and malicious behavior under consideration of the frequencies of system calls. The data set was not published. In [10] older and new algorithmic approaches were compared that evaluate sequences of system calls. It was observed that both the detection and the false alarm rates of the different approaches could not be improved beyond a certain value. The authors' thesis is that the quality of HIDS can be improved if the algorithms also take into account context information such as parameters and return values for system calls. To pursue this thesis, a new data set containing such information is needed.

```
# Structure of KDD BSM data.
open(2): read
system call      open(2)
event-ID         72 AUE_OPEN_R
event class      fr(0x00000001)
audit record:    header token, path token, [attr token], subject token, return token

# Extract from the UNM data set: PID SystemcallID, PID SystemcallID, ...
162 4, 162 2, 162 66, ...

# Extract from the ADFA-LD data set: SystemcallID SystemcallID ...
54 175 120 ...

# Extract from the NGID-DS data set
DATA, TIME, PID, PATH, SystemcallID, Event ID, Categ., Subcat, Label
11/03/2016, 2:45:01, 1830, /sbin/upstart-dbus-bridge, 142, 45354, normal, normal, 0
11/03/2016, 2:45:06, 1804, /bin/dbus-daemon, 256, 45352, normal, normal, 0
```

Listing 1. Fragments of commonly used IDS data sets

### III. REQUIREMENTS

Based on the weaknesses of the previous data sets [8], the following requirements apply to the new data set and the method of producing it: Over time, the number, syntax and semantics of system calls of operating systems have changed. For this reason and to solve the lack of topicality, the system calls of today's systems and current software must be considered. To ensure that the generated data sets can be kept up-to-date in the future, the simulation process should be replicable. To fix the lack of thread information, the recorded system calls must contain process and thread information. This allows the data set to correctly represent normal and attack behavior in today's multithreaded environments. In addition, the recorded system calls must include metadata such as their time stamps, parameter and return values to solve the mentioned lack of meta information. The size of the data set, i.e. the number of contained sequences and their system calls can be selected as required in order to carry out procedures with large training requirements, such as the training of a neural network. This solves the lack of data volume. Normal and attack behavior shall be recorded

<sup>1</sup><https://software.intel.com/en-us/articles/pin-a-dynamic-binary-instrumentation-tool>

according to the same procedure. The basic conditions such as operating system/kernel, the software used and versions should be identical. The only difference between normal and attack procedures is the attack carried out during the simulation. The process must be customizable in order to be able to adapt the collected data to the respective application area by implementing own scenarios.

### IV. METHOD

Our method considers the previously established requirements to collect suitable data sets for training and evaluation of a HIDS. By using this method, scenarios can be defined and both normal behaviour and attack behavior can be simulated and recorded. Simulation in this matter means the staged execution of benign and attacker behavior on an actual machine. The method basically defines the following procedure pattern: (1) The acquisition of events at kernel level (system calls). (2) The use of a container-virtualized environment. (3) A framework for instrumentation and configuration of scenarios.

The procedure is based on a system model in which three actors *Victim Unit*, *Normal Behavior Unit* and *Control Unit* are related. The development of a system model represents an application scenario. The Leipzig Intrusion Detection Data Set (LID-DS) framework was developed and provided as a reference implementation.

With it, scenarios that include or exclude vulnerabilities can be defined, simulated and recorded. In recent years, Microservice architecture has become more and more established. With such an architecture, complex software is composed of many loosely coupled services. Due to this development, the whole process is adapted for use in container virtualized environments. Therefore, the resulting dataset no longer describes an entire complex system but a single component, e. g. a web server. This is referred as a scenario.

The result is a set of captured instructions executed by the system in the form of system calls related to single application. This shifts the scope towards attack vectors based on behaviour and network communication. However, many choices pointed out here only require a little adjustment to support other scenarios.

#### A. System model of an application scenario

Scenarios implemented in the LID-DS are based on software using the client-server model. However, also centralized alternatives like the mainframe architecture or peer-to-peer applications are within the scope of the recording framework. An example is the web server scenario, which consists of one web server as well as  $n$  units which request it. Figure 1 list the schematic model in normal and attack behavior. The *Victim Container* describes the *Server* in the network. This is monitored by a *Container Sensor* that extracts normal and attack behavior. The Sensor works introspectively and records the events on kernel level in the form of system calls with minimal influencing the behaviour. *Normal Behavior Unit(s)* resemble the clients in the network. In a web server scenario, they would execute requests to call a web page. As with reality,

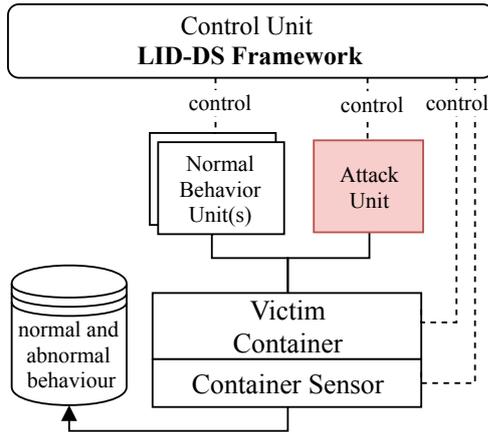


Fig. 1. system architecture to capture attack behavior

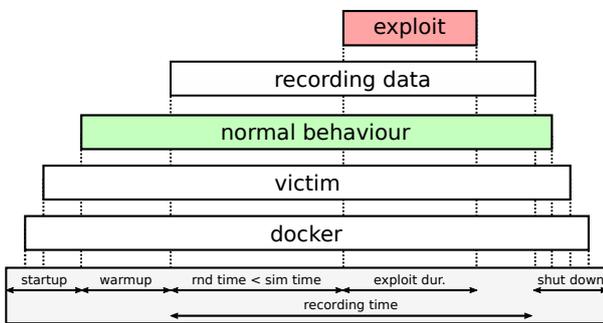


Fig. 2. Simulation procedure of LID-DS

several of these actors can exist to create a realistic normal behavior, including occasional load peaks. The *Attack Unit* also is a client in the network, which however executes an attack on the victim.

Normal behavior and attack behavior are caused by the influence of other entities like containers or processes. In a server-client application for example it is possible to map one or more webclients to a container. A separate *control unit* executes the LID-DS framework and implements the application scenario. LID-DS allows controlling the lifecycle of containers, which includes the initialization and configuration of the actors as well as the control and monitoring of the entire simulation process. Use cases are defined in the form of scenarios and executed by the process.

In this way, simple as well as more complex realistic scenarios, such as multi-step attacks, can be evaluated.

### B. Simulation procedure

The simulation process is shown in figure 2 and consists of the four consecutive phases *startup*, *warmup*, *simulation of normal program behavior*, *shutdown* and potentially the *exploit*. Basically, each implemented scenario is executed according to this procedure. At the beginning, the container is initialized and the Victim Container is configured and executed. At this point, benign user behavior is started to be executed with respect to the victim container. The subsequent

*warmup* phase includes a delay so that the system is in a steady (non-transient) state when recording starts. After the *warmup* delay has passed, the recording of the normal behavior begins. Granted, a malicious user pattern is supplied, the attack behavior gets executed at a randomly chosen time within the recording window. After the specified recording time has passed, the monitoring tool together with all containers gets shut down in reverse order that they started.

### C. Instructional System Call Tracer

Historically, *strace*<sup>2</sup> was used to monitor interactions between processes and the Linux kernel. *Strace* interrupts the traces process every time a system call is invoked, captures the system call, decodes it and then resumes the execution of the monitored process. It is obvious that while this behavior allows for easy recording and tampering of system instructions, for the purpose of recording system activity this is not very efficient. *Sysdig*<sup>3</sup> on the other hand, loads a small driver in the kernel that makes it possible to handle different events related to system calls. This event collection is, in contrast to *strace*, non-blocking. Furthermore, *Sysdig* pre-processes the data collected, combining information on system call executions with data from *tcpdump* or information on referenced files. Our approach settled on using *Sysdig* for recording system calls since it provides a pragmatical way of achieving the requirements identified here. This is best displayed by *Sysdig* providing export functionality, pre-processing of many file descriptors and rich filtering functionality, allowing for efficient prototyping. This choice, however, does not limit the approach's capabilities since *Strace* would allow for the extraction of data in a similar manner.

### D. Container Virtualization Engine

*Docker*<sup>4</sup> as a container virtualization engine (LXC) was chosen because it is a commonly used standard in practice. LXC is used to run multiple instances of the operating system isolated on a single host. In contrast to full virtual machine environments, guests share the kernel with each other. This level of virtualization allows a sufficient isolated environment to be created at the application level. To monitor one or more containers on the host, it is only necessary to inject the Sensor on the host level. This is resource-friendly and allows us to record application behavior with little impact to the system calls. It also gives us a high degree of flexibility in creating scenarios.

### E. LID-DS Framework

The LID-DS framework implements this procedure with minimal development effort. It covers all steps necessary for the simulation and recording of HIDS data. In detail, it takes care of the following steps: handling victim virtualization via *Docker*, System Call Tracing via *Sysdig* and communication between user behavior and victim via a bridge network. To

<sup>2</sup><https://linux.die.net/man/1/strace>

<sup>3</sup><https://github.com/draios/sysdig>

<sup>4</sup><https://www.docker.com>

schedule user actions the distribution introduced by Deng [11] has been chosen.

To record data of a scenario the following information must be defined: A Docker image, specifying the configuration of the victim environment, a set of benign user actions, an script exploiting a vulnerability of the victim and a metric to check for correct and finished initialization of the victim environment.

The LID-DS framework makes it possible to define several normal behaviors for a scenario and the associated victim. All of the passed behaviors are executed in parallel, each in its own thread. This makes it possible to simulate multiple parallel user sessions accessing the victim. This implementation opens up the possibility to mirror real-world network traffic instead of simulating staged user actions to the victim.

Within the scope of real world applications, many different scenarios can hopefully be defined by using a single user simulation definition. Furthermore, many malicious actions, especially actions of reconnaissance are indifferent to many victim configurations. For example, consider a TCP-SYN Scan using the nmap<sup>5</sup> tool.

## V. EVALUATION AND RESULTS

To evaluate how the proposed framework can be used to record host data consider CVE-2012-2122<sup>6</sup>, a tragically comedic security flaw in MariaDB/MySQL. A new data record is created by monitoring a vulnerable MySQL instance according to the procedure shown earlier. The setup consists of a Ubuntu Xenial in version 16.04 with the Docker version 18.09.6 and Sysdig in version 0.24.1. MySQL is used in version 5.5.23, which contains the vulnerability. The simulation time is 5 minutes, the time exploit is 120 seconds. As recording time we have chosen 5 minutes because Sysdig in its report<sup>7</sup> has surveyed an average running time of 5-10 minutes for containers. Two runs are performed. One run only generates normal behavior, a second run contains the attack on the vulnerability in addition to the normal behavior. As a result, the content and technical requirements from III are compared to this data set. The resulting data set is compared to the commonly used IDS data sets on basis of the categorized requirements (Table I).

### A. Comparison

The focus of this comparison is on the resulting artifact, the data set and the features it contains. Concerns related to efficiency during the recording phase are not considered in this paper. The central question is whether this procedure, measured against the result, can solve the criticism and problems of the previous data sets. The evaluation resulted in two data sets. Data set 1 contains normal behavior and includes 46839 system calls. Data set 2 contains normal and attack behavior and includes 128799 system calls. Listing 2 shows an excerpt

<sup>5</sup><https://nmap.org>

<sup>6</sup><https://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2012-2122>

<sup>7</sup><https://sysdig.com/blog/2018-docker-usage-report/>

TABLE I  
FEATURE COMPARISON LID-DS WITH OTHERS

feature	LID-DS	NGIDDS	ADFA-LD	UNM	KDD
topicality	+	+	+	-	-
thread info	+	+	-	+	+
metadata	+	-	-	-	+
data volume	+	+	-	-	+
reproduceability	+	-	-	-	-

of data set 2. For the interpretation of the results, the classified requirements are compared below.

1) *Lack of topicality*: The data was recorded on a modern Linux system, which has over 370 different system calls. The containerized environment and the LID-DS framework makes it easy to repeat such runs for different versions of operating systems. This makes sense because the number of system calls varies from operating system to operating system. The data set can be updated simply by adjusting the configuration and running it again.

2) *lack of thread information*: For each system call in the recording period the thread ID on which the process is running is recorded, as shown in listing 2. This ensures that the multithreading information that is important today is not lost. The most recent data set ADFA-LD lacks this information.

3) *lack of metadata*: For each system call, the data record contains extensive meta information such as high-precision time stamps, process name, transfer parameters and a section of the data buffer. The time stamps in the NGIDDS are only accurate to the second, which can lead to errors in the sequence.

4) *lack of volume*: Over 100 000 system calls were recorded during the survey period. For example, ADFA-LD, as well as the highly obsolete UNM data set, provides a smaller, fixed data set of system calls. The data collection period can be configured in LID-DS so that larger or smaller data records can be generated according to individual requirements.

5) *lack of reproduceability*: By using LXC, the entire simulation can be defined with the LID-DS framework and stored. Including software versions, parameters or configurations. The performed evaluation is stored in Github as *example*<sup>8</sup> and can be viewed and performed by anyone. The framework itself is published under the GNU General Public License. As we know, this is the first time that a process for generating HIDS datasets is available to the public in a fully reproducible form.

### B. Results

Overall, there is a significant superiority of the approach to generating modern HIDS data sets, as shown in table I. By using the LID-DS Framework, all technical and content requirements are fulfilled. The framework allows modern and operating system specific data sets to be generated, which is important to avoid training neural networks on the basis of outdated or incorrect system calls or nowadays uncommon

<sup>8</sup><https://github.com/LID-DS/LID-DS>

```

TIME CPU PROCESS PROCESS_ID ENTER(>)/EXIT(<) SYSCALL ARGUMENTS
t 0 0 apache2 25426 > open
t 1 0 apache2 25426 < open fd=13(<f>/etc/apache2/.htpasswd) name=/etc/apache2/.
      httpasswd flags=4097(O_RDONLY|O_CLOEXEC) mode=0
t 2 0 apache2 25426 > fstat fd=13(<f>/etc/apache2/.htpasswd)
t 3 0 apache2 25426 < fstat res=0
t 4 0 apache2 25426 > read fd=13(<f>/etc/apache2/.htpasswd) size=4096
t 5 0 apache2 25426 < read res=91 data=QUEU75:$apr1$X0JgPVeW$xCKOGdUp2tLNns0t6RqB...
t 6 0 apache2 25426 > close fd=13(<f>/etc/apache2/.htpasswd)
t 7 0 apache2 25426 < close res=0

```

Listing 2. Short excerpt of data from a recorded trace collected with the LID-DS Framework including thread information and metadata

TABLE II  
IMPLEMENTED AND PUBLISHED SCENARIOS BY LID-DS

Scenario	CVE / CWE
Heartbleed	CVE-2014-0160
PHP file upload	CWE-434
Bruteforce login	CWE-307
Rails Disclosure of content	CVE-2019-5418
ZipSlip	various
EPS file upload	CWE-434
MySQL auth bypass	CVE-2012-2122
Nginx int. overflow	CVE-2017-7529
Sprockets info. leak	CVE-2018-3760
SQL injection with sqlmap	CWE-89

operating systems. The amount of data is also important for neural networks, which can be adapted by LID-DS. Multi-threaded information is valuable to view the behavior of a system down to the application and thread level. In this context, the metadata and parameters are also relevant. These can also contain application-specific information and support the correct interpretation of the application behavior. The overall approach provides the basis to effectively compare and evaluate HIDS in the future and to develop new classification features based on thread information, metadata and parameters in order to significantly increase the recognition rate and accuracy of HIDS.

## VI. CONCLUSION

The need for modern, uniform and metadata enhanced data sets can be satisfied by implementing this approach. This way, LID-DS is a significant contribution to future research, evaluation and comparability of Host-Based Intrusion Detection Systems. Additionally, this approach only needs slight adaptations to be functional in production environments. The major advantage of this approach is that it provides a high degree of flexibility in the form of scenarios that can be adapted to individual technical as well as policy requirements. For example, the evaluation MySQL example<sup>9</sup> from chapter V can easily be adapted using real network data. With this approach we have created a new data set. It was published as "Leipzig Intrusion Detection - Data Set (LID-DS)" in [8] which contains different use cases shown in table II. LID-DS framework and ready to use data sets are free to use and published on GitHub<sup>9</sup>. It is the first HIDS data set which contains normal and abnormal behavior, system calls and their timestamps, thread ids, process names,

<sup>9</sup><https://github.com/LID-DS/LID-DS>

arguments, return values and excerpts of their data buffers from traces of normal and attack behavior of several recent, multi-process, multi-threaded scenarios. Many of the included features cannot be extracted from previous data sets. With it, known algorithms can be enhanced or new algorithms, based on the various included features, can be explored. Additionally, staged scenarios based on internal expert knowledge can give a practical prediction on the performance of an algorithm. The approach outlined in this work focuses on giving every actor the possibility to build their own model from their own experienced traffic. Further information on the development of LID-DS and initial analyses can be found in the works of [12] and [13]. An extension of the procedure to include network sensors is planned. Furthermore, an updated version of the data set is scheduled to be released once a year. Every version should expand the data set by including recordings of the latest commonly used software systems as well as disclosed vulnerabilities. Additionally, recordings based on new versions of the Linux kernel and configurations according to new techniques used in development, hosting and pentesting will be the target of these extensions. We anticipate feedback to allow for the progressive development of a data set that finally allows for reproducible IDS research.

## REFERENCES

- [1] M. Pendleton and S. Xu. A dataset generator for next generation system call host intrusion detection systems. In *Proceedings - IEEE Military Communications Conference MILCOM*, volume 2017-Octob, 2017. DOI: 10.1109/MILCOM.2017.8170835.
- [2] Lincoln Laboratory; MIT. DARPA Intrusion Detection Evaluation Data Set. <https://www.ll.mit.edu/r-d/datasets>, 1998-2000.
- [3] Computer Science Department Farris Engineering Center; University of New Mexico. Computer Immune Systems - Data Sets and Software. <https://www.cs.unm.edu/immsec/systemcalls.htm>, 1999.
- [4] C. Warrender, S. Forrest, and B. Pearlmutter. Detecting intrusions using system calls: Alternative data models. In *Proceedings - IEEE Symposium on Security and Privacy*, 1999. DOI: 10.1109/SECPRI.1999.766910.
- [5] Australian Center for Cyber Security (ACCS). The ADFA Intrusion Detection Datasets. <https://www.unsw.adfa.edu.au/australian-centre-for-cyber-security/cybersecurity/ADFA-IDS-Datasets/>, 2013.
- [6] G. Creech and J. Hu. Generation of a new IDS test dataset: Time to retire the KDD collection. In *IEEE Wireless Communications and Networking Conference, WCNC*, 2013. DOI: 10.1109/WCNC.2013.6555301.
- [7] W. Haider, J. Hu, J. Slay, B.P. Turnbull, and Y. Xie. Generating realistic intrusion detection system dataset based on fuzzy qualitative modeling. *Journal of Network and Computer Applications*, 87:185–192, 6 2017. DOI: 10.1016/J.JNCA.2017.03.018.
- [8] M. Grimmer, M. M. Röhlung, D. Kreusel, and S. Ganz. A modern and sophisticated host based intrusion detection data set. In *IT-Sicherheit als Voraussetzung für eine erfolgreiche Digitalisierung*, pages 135–145, 2019. ISBN: 978-3-922746-82-9.
- [9] A. S. Abed, C. Clancy, and D. S. Levy. Intrusion detection system for applications using linux containers. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 9331, pages 123–135, 11 2015. DOI: 10.1007/978-3-319-24858-5\_8.
- [10] M. Grimmer, M. M. Röhlung, M. Kricke, B. Franczyk, and E. Rahm. Intrusion Detection on System Call Graphs. In *Sicherheit in vernetzten Systemen*, pages G1–G18, 2018. ISBN: 978-3-3-7460-8637-8.
- [11] Deng, S. Empirical model of WWW document arrivals at access link. In *Proceedings of ICC/SUPERCOMM '96 - International Conference on Communications*, volume 3, pages 1797–1802. IEEE. DOI: 10.1109/ICC.1996.535600.
- [12] S. Ganz. Ein moderner Host Intrusion Detection Datensatz, 2019.
- [13] D. Kreußel. Simulation and analysis of system call traces for adversarial anomaly detection, 2019.



# 3<sup>rd</sup> Workshop on Internet of Things—Enablers, Challenges and Applications

**T**HE Internet of Things is a technology which is rapidly emerging the world. IoT applications include: smart city initiatives, wearable devices aimed to real-time health monitoring, smart homes and buildings, smart vehicles, environment monitoring, intelligent border protection, logistics support. The Internet of Things is a paradigm that assumes a pervasive presence in the environment of many smart things, including sensors, actuators, embedded systems and other similar devices. Widespread connectivity, getting cheaper smart devices and a great demand for data, testify to that the IoT will continue to grow by leaps and bounds. The business models of various industries are being redesigned on basis of the IoT paradigm. But the successful deployment of the IoT is conditioned by the progress in solving many problems. These issues are as the following:

- The integration of heterogeneous sensors and systems with different technologies taking account environmental constraints, and data confidentiality levels;
- Big challenges on information management for the applications of IoT in different fields (trustworthiness, provenance, privacy);
- Security challenges related to co-existence and interconnection of many IoT networks;
- Challenges related to reliability and dependability, especially when the IoT becomes the mission critical component;
- Zero-configuration or other convenient approaches to simplify the deployment and configuration of IoT and self-healing of IoT networks;
- Knowledge discovery, especially semantic and syntactical discovering of the information from data provided by IoT;

The IoT conference is seeking original, high quality research papers related to such topics. The conference will also solicit papers about current implementation efforts, research results, as well as position statements from industry and academia regarding applications of IoT. The focus areas will be, but not limited to, the challenges on networking and information management, security and ensuring privacy, logistics, situation awareness, and medical care.

## TOPICS

The IoT conference is seeking original, high quality research papers related to following topics:

- Future communication technologies (Future Internet; Wireless Sensor Networks; Web-services, 5G, 4G, LTE, LTE-Advanced; WLAN, WPAN; Small cell Networks...) for IoT,

- Intelligent Internet Communication,
- IoT Standards,
- Networking Technologies for IoT,
- Protocols and Algorithms for IoT,
- Self-Organization and Self-Healing of IoT Networks,
- Trust, Identity Management and Object Recognition,
- Object Naming, Security and Privacy in the IoT Environment,
- Security Issues of IoT,
- Integration of Heterogeneous Networks, Sensors and Systems,
- Context Modeling, Reasoning and Context-aware Computing,
- Fault-Tolerant Networking for Content Dissemination,
- Architecture Design, Interoperability and Technologies,
- Data or Power Management for IoT,
- Fog—Cloud Interactions and Enabling Protocols,
- Reliability and Dependability of mission critical IoT,
- Unmanned-Aerial-Vehicles (UAV) Platforms, Swarms and Networking,
- Data Analytics for IoT,
- Artificial Intelligence and IoT,
- Applications of IoT (Healthcare, Military, Logistics, Supply Chains, Agriculture, ...),
- E-commerce and IoT.

The conference will also solicit papers about current implementation efforts, research results, as well as position statements from industry and academia regarding applications of IoT. Focus areas will be, but not limited to above mentioned topics.

## EVENT CHAIRS

- **Cao, Ning**, College of Information Engineering, Qingdao Binhai University
- **Furtak, Janusz**, Military University of Technology, Poland
- **Hodoň, Michal**, University of Žilina, Slovakia
- **Zieliński, Zbigniew**, Military University of Technology, Poland

## PROGRAM COMMITTEE

- **Al-Anbuky, Adnan**, Auckland University of Technology, New Zealand
- **Antkiewicz, Ryszard**, Military University of Technology, Poland
- **Baranov, Alexander**, Russian State University of Aviation Technology, Russia

- **Brida, Peter**, University of Zilina, Slovakia
- **Chudzikiewicz, Jan**, Military University of Technology in Warsaw, Poland
- **Cui, Huanqing**, Shandong University of Science and Technology, China
- **Dadarlat, Vasile-Teodor**, Univerversita Tehnica Cluj-Napoca, Romania
- **Ding, Jianrui**, Harbin Institute of Technology, China
- **Diviš, Zdenek**, VŠB-TU Ostrava, Czech Republic
- **Fortino, Giancarlo**, Università della Calabria
- **Fouchal, Hacene**, University of Reims Champagne-Ardenne, France
- **Fuchs, Christoph**, Fraunhofer Institute for Communication, Information Processing and Ergonomics FKIE, Germany
- **Giusti, Alessandro**, CyRIC - Cyprus Research and Innovation Center, Cyprus
- **Hudík, Martin**, University of Zilina
- **Husár, Peter**, Technische Universität Ilmenau, Germany
- **Johnsen, Frank T.**, Norwegian Defence Research Establishment (FFI), Norway
- **Jurecka, Matus**, University of Žilina, Slovakia
- **Kafetzoglou, Stella**, National Technical University of Athens, Greece
- **Kapitulík, Ján**, University of Žilina, Slovakia
- **Karastoyanov, Dimitar**, Bulgarian Academy of Sciences, Bulgaria
- **Karpiš, Ondrej**, University of Žilina, Slovakia
- **Kochláň, Michal**, University of Žilina, Slovakia
- **Krco, Srdjan**, DunavNET
- **Laqua, Daniel**, Technische Universität Ilmenau, Germany
- **Lenk, Peter**, NATO Communications and Information Agency, Other
- **Li, Guofu**, University of Shanghai for Science and Technology, China
- **Marks, Michał**, NASK - Research and Academic Computer Network, Poland
- **Milanová, Jana**, University of Žilina, Slovakia
- **Monov, Vladimir V.**, Bulgarian Academy of Sciences, Bulgaria
- **Murawski, Krzysztof**, Military University of Technology, Poland
- **Niewiadomska-Szynkiewicz, Ewa**, Research and Academic Computer Network (NASK), Institute of Control and Computation Engineering, Warsaw University of Technology
- **Ohashi, Masayoshi**, Advanced Telecommunications Research Institute International / Fukuoka University, Japan
- **Papaj, Jan**, Technical university of Košice, Slovakia
- **Ramadan, Rabie**, Cairo University, Egypt
- **Ševčík, Peter**, University of Žilina, Slovakia
- **Shaaban, Eman**, Ain-Shams university, Egypt
- **Shu, Lei**, Guangdong University of Petrochemical Technology, China
- **Skarmeta, Antonio**, University of Murcia
- **Smirnov, Alexander**, Linux-WSN, Linux Based Wireless Sensor Networks, Russia
- **Staub, Thomas**, Data Fusion Research Center (DFRC) AG, Switzerland
- **Suri, Niranjana**, Institute of Human and Machine Cognition
- **Teslyuk, Vasyl**, Lviv Polytechnic National University, Ukraine
- **Wang, Zhonglei**, Karlsruhe Institute of Technology, Germany
- **Wrona, Konrad**, NATO Communications and Information Agency
- **Xiao, Yang**, The University of Alabama, United States
- **Zhang, Tengfei**, Nanjing University of Post and Telecommunication, China

# Remote Programming and Reconfiguration System for Embedded Devices

Tomasz Michalec, Maksymilian Wojczuk, Robert Brzoza-Woch, Tomasz Szydło  
AGH University of Science and Technology,  
Department of Computer Science, Krakow, Poland.  
Email: robert.brzoza@agh.edu.pl

**Abstract**—This article presents a concept of a system which can be utilized as a remote management add-on for embedded devices. It can be applied to resource-constrained wireless sensors and IoT nodes based on a general purpose microcontroller unit or a field programmable gate array (FPGA) chip. The proposed solution facilitates remote firmware update, management, and operation monitoring. Thanks to the utilization of standard protocols and interfaces, the proposed system is very flexible and it can be easily customized for multiple modern microcontrollers or programmable logic chips. The presented system can be an efficient solution for fast prototyping and it can be an alternative to a time-consuming process of bootloader development for ad hoc devices. It can also be applied to remote laboratory access for educational purposes. A proof of concept prototype implementation has been successfully developed and evaluated. The implementation is available on a free license and utilizes a commonly available and inexpensive hardware platform.

## I. INTRODUCTION

INTERNET of Things (IoT) uses multiple nodes distributed among different physical locations. The nodes may require remote management and firmware upgrade mechanisms to be implemented. As the IoT systems are often utilized for monitoring and interaction with an environment, their operation has to be either well simulated or tested in a laboratory or in a target environment. If embedded or IoT software developers choose the approach that involves practical testing, the problem of the remote management of IoT nodes arises – it includes monitoring of a node operating condition, setting its operation parameters, and upgrading its firmware.

In production environments, a common approach of deploying a remotely manageable embedded device is to implement a bootloader. Unfortunately, the process of developing a bootloader software may be a very demanding and complex task – the software needs to be well tested because it is a crucial part of the system. In case of the bootloader malfunction, while the device's software development is at an early stage, the device becomes unusable until reprogrammed directly through a local interface. That requires physical access to the device's hardware – it can be very inconvenient in the domain of IoT and sensor nodes which may operate in remote locations.

In this paper, we present a concept and a sample implementation of a versatile add-on subsystem for remote management of embedded devices, IoT platforms, especially based on resource-constrained MCUs.

## II. RELATED RESEARCH AND AVAILABLE SOLUTIONS

A common scientific issue is designing a remote laboratory which is usually utilized to allow for remote access to laboratory infrastructure via the Internet [1], [2]. There are presented extensions of this concept, which allow designers to implement the remote laboratory on a single-board computer (SBC), but still it requires to run a full-featured operating system, e.g. Linux [3]. Such an operating system requires a large amount of hardware resources and energy.

A natural solution for firmware updates is to write a bootloader program. However, such a bootloader must be extremely reliable and it is more difficult to write a reliable bootloader which itself could be updated remotely using e.g. wireless connection [4]. Currently, one of the leading solutions in the field of remote management of the embedded devices is the utilization of the OMA Lightweight M2M (LWM2M) protocol [5], [6]. It is based on Constrained Application Protocol (CoAP) which is popular in the IoT domain [7] due to its relatively low resource requirements.

There are commercial solutions requiring additional hardware that could program flash memories of MCUs through the network, e.g. XDS220 USB/Ethernet JTAG Emulator or the Intel FPGA Ethernet Cable as used in [8]. However, those solutions are usually expensive, and their application is usually limited to a vendor-dependent subset of supported devices. Considering their application for a large number of managed nodes might not be economical.

A time-efficient approach for remote programming, software development and prototyping of embedded devices [9] may be more convenient if the remote reconfiguration is applied. The utilization of SBCs, such as Raspberry Pi, becomes more and more popular [10] even for high-end and military applications [11]. When equipped with proper software, such as the OpenOCD [12], [13], the Raspberry Pi SBCs can become remote management nodes as in e.g. [14].

## III. PROBLEM STATEMENT

After analyzing the available literature and solutions we decided to develop a concept of the remote development tool for MCU-based embedded systems with an option to expand its functionality to remote reconfiguration of FPGAs. The concept should allow designers to develop, customize, and deploy the versatile remote programmer-monitor tool for facilitating embedded software development.

The discussed problem concerns the development of multi-node systems based on embedded devices that require remote and batch firmware updates. The designed solution should meet the following crucial requirements: (1) the solution should be versatile or at least easy to adapt and extend for various MCU hardware platforms with an option for future FPGA configuration support; (2) the remote programming system should allow for easy interaction with remote embedded devices – mainly the firmware update; (3) it should detect the connected target board and adjust parameters automatically; (4) The remote reconfiguration tool should be able to operate on a commonly available and inexpensive hardware platform; (5) the interface for the user or a developer should be platform-independent.

#### IV. THE DESIGN CONCEPTS

In this section, we present suggested choices and concepts for implementing the remote reconfiguration system based on the requirements stated in Section III.

##### A. General architecture and communication

We propose the following architecture of the remote programming system. The system may consist of two separate main parts: (1) the hardware part called the Remote Programming Device (RPD) further in this article and (2) the management part which is a user application for interacting with one or more RPDs.

The *RPD* is able to remotely reprogram internal memories of microcontrollers with provided binary firmware files and optionally to reconfigure FPGA integrated circuits. Primarily is intended to work as a temporary add-on to an embedded device or an IoT node during final stages of software development. It can also be utilized for diagnostic and long-term monitoring purposes in prototype and experimental IoT systems. The embedded device, which is managed, reprogrammed, or reconfigured by the RPD, is in this paper referred to as the *target device*.

The *management part* is intended to run on a user's host computer. For the purposes of communication between the two parts of the system, we have chosen and recommend the LWM2M protocol due to its popularity, basic security, and ability to communicate not only within local networks but also globally over the Internet. As the management part, we decided to use the Eclipse Leshan<sup>1</sup> implementation of the LWM2M server. It provides a Web-based user interface (UI) which allows for interaction with connected RPDs. That interface does not need any additional specific software to be installed on the management computer.

To provide versatility, the RPD is recommended to communicate with the target device using a standard and popular interface, primarily Joint Test Action Group IEEE 1149.1 (JTAG) and, eventually, Serial Wire Debug (SWD). An MCU for RPD may be a typical inexpensive unit for embedded systems purposes. The Ethernet was chosen to implement a

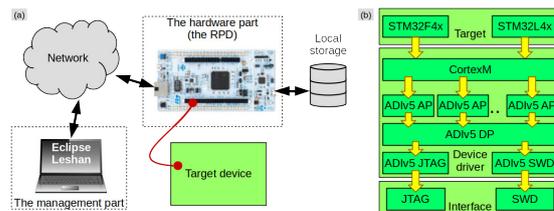


Fig. 1. RPD architecture and usage (a) and RPD abstraction layers diagram (b).

convenient physical layer for Internet Protocol (IP) communication. For the prototype implementation, we have chosen the Nucleo boards equipped with STM32F429ZI MCU with ARM Cortex-M4 microprocessor core. RPD uses a USB flash drive to store programming files and configuration locally.

The proposed remote programming architecture has been shown in Figure 1a.

#### V. SELECTED DETAILS OF SAMPLE IMPLEMENTATION

A sample software for the remote programming system has been successfully implemented. This section contains selected details which concern practical aspects of the remote programming system operation.

##### A. RPD general architecture

The RPD has been designed to be easily extendable. Its architecture is based on layers. Figure 1b shows the designed organization of the layers. The first layer, denoted as *Target*, is the only layer exposed to the external interface described further in Section V-B. The *Target* layer purpose is to represent all programmable devices in a unified way. The *Device driver* is an intermediate layer which can be partitioned into multiple sub-layers. The driver sub-layers are able to communicate with devices supporting ARM Debug Interface Access and Data Ports (ADiv5 AP and ADiv5 DP). Such an architecture allows similar devices to share common parts of software implementation. This layer is independent of hardware. The *Interface* is the lowest layer and it encapsulates logic needed to use a physical medium. It is tightly coupled with hardware used to realize the RPD.

As a proof of concept implementation, we provide support for programming two different MCUs. The users and developers can extend the range of the supported programmed chips by modifying the provided source code. Further details on the RPD are elaborated in Section V-C.

##### B. Communication part

We used a standard LWM2M protocol stack with the CoAP over User Datagram Protocol (UDP) implemented with the LwIP stack – a commonly used TCP/IP stack designed for embedded devices. We have used Eclipse Leshan as the LWM2M server to provide the user with a generic Web-based UI for managing the programmer resources.

The communication with the management part includes two parts: the management interface using the LWM2M and the file download part with the Hypertext Transfer Protocol (HTTP). The LWM2M implementation at the RPD side uses

<sup>1</sup><https://www.eclipse.org/leshan/>

the Wakaama code [5], [15]. The user can browse, read and update each connected device's properties through the UI. LWM2M server sends user's actions to particular devices and calls adequate procedures associated with resources. The firmware URL resource contains the URL of the current binary file. Once updated, the device downloads the newest firmware version from the HTTP server. A simplified process has been shown in Figure 2.

The whole communication between the user and the RPD is done through the LWM2M Server, excluding the binary download which is done using the HTTP. In order to ensure basic functionality for the sample implementation, we have defined an object representing the Remote Target – an embedded device to be programmed. The object contains properties necessary to monitor and control vital aspects of the remote target. The network configuration mechanism has been implemented and it can be dynamically changed without direct physical access to the device.

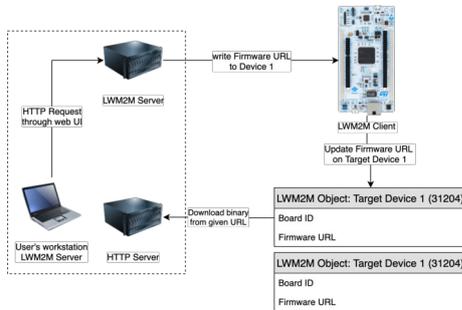


Fig. 2. Simplified communication flow (the LWM2M Server, the HTTP Servers, and the user's workstation may or may not be running on a single machine).

### C. Programming part

The sample implementation of the RPD is able to program flash memory of the following MCU families: STM32F4xx, STM32L4xx. Both of them are similar, but the STM32L4xx uses more energy-efficient technology and has updated hardware peripheral modules. JTAG has been selected as the hardware interface for the MCU programming in the prototype implementation. The JTAG's daisy-chaining feature is supported in the sample firmware. The RPD is also able to perform automatic discovery of connected devices by using their IDCODE registers. Discovered devices are exposed as independent targets to the management part.

## VI. PROTOTYPE EVALUATION

Further in this section, we present a quantitative evaluation of the sample implementation of the RPD and its analysis.

### A. Network communication

The system presented in this paper is intended to provide the possibility to transfer a new firmware binary file from the *management part* to a target embedded device programmed by an RPD. The main goal of the system is to allow for programming the target located in a distant physical location. To prove that functionality and usefulness of the created

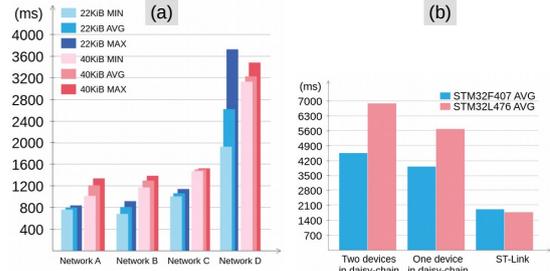


Fig. 3. Comparison of binary files download time (a) and programming time (b).

system we have tested time performance for different network environments. We conducted 5 practical tests for each of the following network conditions:

- **Network A:** The LWM2M server, the HTTP server, the user's computer, and the RPD with target boards are located in the same local network. The user's computer is connected with the router and switch through Wi-Fi interface.
- **Network B:** The LWM2M server, the user's computer, and the RPD are placed in the same local network, but the binary file is downloaded from an external HTTP server in another network but in the same city Kraków, Poland.
- **Network C:** The LWM2M server is running on the Amazon Elastic Compute Cloud (Amazon EC2) instance from Amazon Web Services in Frankfurt Datacenter and the binary files are placed also in Frankfurt, on the Amazon Simple Storage Service (the Amazon S3). User's computer, the RPD connected to target boards are located in Kraków, Poland, in the same local network.
- **Network D:** In this test set-up, the user's computer as a virtual machine, the LWM2M Server, and the HTTP Server on the EC2 Instance from Amazon Web Services were located in the United States Datacenter while the RPD was in the AGH University network in Kraków, Poland, Europe.

The results are shown in Figure 3a. The binary download time to the target is similar for the local network conditions and across neighbouring countries on one continent – in all of those cases the system has a similar level of its overall usefulness and the physical distance had only a limited impact on the overall system performance. The network overhead plays a greater role in large distances as in the *Network D* case. However, system can then still be considered useful because the binary download time does not exceed 4 s.

### B. The target device programming

This section discusses its overall performance in different scenarios with reference to the underlying dependencies of this process. The information can be useful in comparison with other available commercial solutions than mentioned in this section and also to provide more complete information for scientists and engineers who wish to contribute to the RPD development.

Time required for the programming process includes overhead for communication between the RPD and a target device,

erasing the target's flash memory, and the target's flash memory programming.

The JTAG clock signal (TCK) was set to 1 MHz. For erasing flash memory on STM32F4xx and STM32L4xx we utilized a mechanism to erase only those memory regions that were going to be programmed. For programming the STM32L4xx we used a binary file with size 40 KiB. STM32F4xx was programmed using a binary file with size of 22 KiB.

Tests in the following scenarios have been performed: (1) RPD with two devices in JTAG daisy-chain, (2) RPD with one device in JTAG chain, and (3) ST-Link programmer connected locally with USB. The latter scenario served as a reference for comparison with a commercial solution. All tests were repeated 5 times and the average time is presented in Figure 3b.

The process of programming a target device while another one is in JTAG daisy-chain takes longer than programming one target device only, because it was required to write the *BYPASS* instruction to all other devices in the chain. There is also additional code in the RPD that needs to be executed to handle multiple targets.

Even the STM32L4xx binary file is almost two times larger than the STM32F4xx program, time for flashing the STM32L4xx increased only 145%, because it also includes an overhead for programming operation.

The RPD firmware was compiled with the free GCC cross compiler, arm-none-eabi-gcc. With the compiler optimization level set to *Og*, the resulting firmware size is 149.2 KiB and the RPD program requires 135.4 KiB of static data memory.

The measured power consumption of the RPD is 1.1 W with fully operational Ethernet interface but without an external mass storage device. The total power consumption may vary depending on the utilized mass storage memory, usually a USB flash drive. In practice, the measured total power consumption with a flash drive connected is approx. 1.2 W.

## VII. CONCLUSION AND FUTURE WORK

In this article, we present a concept of the remote programming, configuration, and monitoring system for development and testing of embedded devices and IoT nodes.

The sample implementation of the reconfiguration system has been successful, the basic proof-of-concept functionality has been achieved, and the requirements stated in Section III are met which proves the overall correctness of the presented concept. The remote management interface is easy to use and can be run on many different operating systems thanks to the utilization of the Leshan LWM2M implementation with Web-based GUI. The created RPD software is ready for users to further develop the RPD functionality according to their own use cases, including new programmed devices, sensors, etc. The developed software is available on GitHub<sup>2</sup> on the free (MIT) license. The hardware price of the remote programming system presented in this paper is much lower than the commercial solutions presented in Section II. However, this implementation may require additional work needed for

customizing it for specific, not yet supported use cases. The additional labor cost can be less noticeable if the multiple RPDs with the same custom firmware are deployed.

Future improvements may include implementing a support for the SWD interface, optimizing storage management and flash memory writing, as well as remote management of the RPD firmware and health checks of the connected target devices using additional sensors.

## ACKNOWLEDGMENT

The research presented in this paper was partially supported by the National Centre for Research and Development (NCBiR) under Grant No. LIDER/15/0144/L-7/15/NCBR/2016.

## REFERENCES

- [1] R. Bose, "Virtual labs project: A paradigm shift in internet-based remote experimentation," *IEEE access*, vol. 1, pp. 718–725, 2013. [Online]. Available: <https://doi.org/10.1109/ACCESS.2013.2286202>
- [2] A. V. Parkhomenko, O. Gladkova, E. Ivanov, A. Sokolyanskii, and S. Kurson, "Development and application of remote laboratory for embedded systems design," *International Journal of Online Engineering (iJOE)*, vol. 11, no. 3, pp. 27–31, 2015.
- [3] P. Alexander and N. Radhakrishnan, "Remote lab implementation on an embedded web server," in *2015 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2015]*. IEEE, 2015, pp. 1–5. [Online]. Available: <https://doi.org/10.1109/ICCPCT.2015.7159525>
- [4] S. Schmidt, M. Tausig, M. Hudler, and G. Simhandl, "Secure firmware update over the air in the internet of things focusing on flexibility and feasibility," in *Internet of Things Software Update Workshop (IoTSU) Proceeding*, 2016.
- [5] S. Rao, D. Chendanda, C. Deshpande, and V. Lakkundi, "Implementing LWM2M in constrained IoT devices," in *2015 IEEE Conference on Wireless Sensors (ICWiSe)*. IEEE, 2015, pp. 52–57. [Online]. Available: <https://doi.org/10.1109/ICWiSe.2015.7380353>
- [6] J. Prado, "OMA Lightweight M2M Resource Model," in *IAB IoT Semantic Interoperability Workshop*, 2016.
- [7] B. Djamaa, M. A. Kouda, A. Yachir, and T. Kenaza, "Fetchiot: Efficient resource fetching for the internet of things," in *2018 Federated Conference on Computer Science and Information Systems (FedCSIS)*. IEEE, 2018, pp. 637–643. [Online]. Available: <http://dx.doi.org/10.15439/978-83-949419-5-6>
- [8] J. Belleman, D. Belohrad, L. Jensen, M. Krupa, and A. Topaloudis, "The LHC Fast Beam Current Change Monitor," *WEPF29, IBIC*, 2013.
- [9] A. Tutaj and J. Augustyn, "Universal serial bus as a communication medium for prototype networked data acquisition and control systems-performance optimisation and evaluation," in *2018 Federated Conference on Computer Science and Information Systems (FedCSIS)*. IEEE, 2018, pp. 665–674. [Online]. Available: <http://dx.doi.org/10.15439/978-83-949419-5-6>
- [10] R. Baumgartl and D. Muller, "Raspberry pi as an inexpensive platform for real-time traffic jam analysis on the road," in *2018 Federated Conference on Computer Science and Information Systems (FedCSIS)*. IEEE, 2018, pp. 623–627. [Online]. Available: <http://dx.doi.org/10.15439/978-83-949419-5-6>
- [11] F. T. Johnsen, "Using publish/subscribe for short-lived iot data," in *2018 Federated Conference on Computer Science and Information Systems (FedCSIS)*. IEEE, 2018, pp. 645–649. [Online]. Available: <http://dx.doi.org/10.15439/978-83-949419-5-6>
- [12] H. Högl and D. Rath, "Open on-chip debugger–openocd-," *Fakultat für Informatik, Tech. Rep.*, 2006.
- [13] D. Rath, "Openocd," <https://github.com/ntfreak/openocd>, 2005.
- [14] R. Brzoza-Woch, Ł. Gurdek, and T. Szydło, "Rapid embedded systems prototyping—an effective approach to embedded systems development," in *2018 Federated Conference on Computer Science and Information Systems (FedCSIS)*. IEEE, 2018, pp. 629–636. [Online]. Available: <http://dx.doi.org/10.15439/978-83-949419-5-6>
- [15] O. M. Alliance, "Lwm2m specification 1.0," *Open Mobile Alliance: San Diego, CA, USA*, 2017.

<sup>2</sup><https://github.com/maxiwjoj/RemoteProgrammer>

# A Framework for Autonomous UAV Swarm Behavior Simulation

Piotr Cybulski

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

**Abstract**—In the last several years a large interest in the unmanned aerial vehicles (UAVs) has been seen. This is mostly due to an increase of computational power and decreasing cost of the UAVs itself. One of an intensively researched area is an application of a swarm behavior within team of such UAVs. Simulation tools are one of the means with which quality of solutions in this matter can be measured. In this paper such simulation framework is proposed. The proposed framework is capable of taking under consideration interferences between communicating UAVs, as well as interaction between UAV and surrounding environment. Mathematical models based on which simulation is performed were described, definition of simulation scenario and results of exemplary simulation were also presented.

## I. INTRODUCTION

THE unmanned aerial vehicles, commonly called as drones, are gaining more interest by both civilian and military organizations. From an academic perspective drones are specially interesting because of the swarm intelligence, that can be implemented into them. Combining the artificial intelligence (AI) with the UAV swarms can significantly change the way of providing services such as traffic monitoring, area patrolling or, in the military area of interests, creating a situational awareness. There are many propositions how to approach the swarm intelligence subject so that results will meet the requirements [1], but there is a lack of a common framework to compare results. One of the reasons is large variety of task types for the swarms. To name a few, search and attack on a target[2][3][4], area patrolling[5][6], disaster operations[7] or transportation services[8]. Currently, there are tools for simulating specific algorithms[9] or certain scenarios [7]. There is, however, lack of a general purpose tool for UAV swarm simulation. In this paper I presented the simulation framework to fill this gap.

## II. RELATED WORK

Due to large interest in the area of drone swarms, there have been developed few tools for simulating them. These are mainly for prototyping solutions and studying how does single UAV, or swarm of them, act in a particular scenario. Majority of them is designed to simulate a predetermined type of scenarios such as search and attack on target[4][9] or disaster operations[7]. Some of currently working simulators are based

on commercial software such as Matlab[10] or X-Plane[11]. Finally, there are programming languages, the Proto[12] for example, designed specifically for testing certain paradigms (amorphous computing in this case) application in the UAV swarms area.

## III. PRELIMINARIES AND BASIC DEFINITIONS

Using the definition from [1] let's define the UAV swarm as a team of an autonomous unmanned aerial vehicles, where behavior of a single UAV emerges from its inner state and from surrounding environment, including neighbor UAVs. Because of wide range of tasks for UAV swarms and scale of solutions for these tasks, the following assumptions were made regarding algorithms which are compatible with proposed framework:

- 1) The UAV swarm is controllable in decentralized manner;
- 2) No common knowledge database is being used. Each UAV has to either have necessary information, or it has to be able to obtain it during a mission;
- 3) An environment in which swarm is operating is continuous three dimensional space, or it must be convertible to such.

Each simulation object “inside” framework is treated in the same way in terms of controlling its behavior. It is possible to combine the fully autonomous object, such as swarm's members, with manually controlled units (e.g. proxies between an operator and the swarm). The later described framework is not restricted to any particular types of a mission. However, it is important to point out that mission objectives supported by the framework must be evaluable solely on the situation in the moment at which evaluation is performed. We will define mission objectives as a set of states simulation objects must be in. Formally a mission objective will be described in IV-C

## IV. THE BASIC SWARM MODEL

The main inspiration for basic swarm model was taken from “the asynchronous event-driven robotic network” presented in [13]. Following are main differences between basic swarm model and the one mentioned above:

- 1) A sender of a message is not known explicitly to a receiver, they can be inferred (if there is a need for it) from the message itself;

- 2) In each simulation moment there can be generated multiple messages by all objects.

#### A. The goal

The basic swarm model shall be capable of simulating communication between objects. Communication, if occurs, is not altered in any way between a sender and a receiver. In other words, interference is not taken under consideration.

#### B. The description of basic model's components

A core element of the model is a *simulation object*. Every entity modeled within simulation has to be considered as a simulation object. There are two types of simulation objects: passive and active. Passive simulation objects are aimed to represent entities such as:

- 1) On-the-ground beacons;
- 2) The GPS;
- 3) An environment.

These objects play important role in the behavior of the swarm (and not only that), but their key trait is that they are not changing its behavior due to an interaction with "outside world" (other simulation objects). Active simulation objects on the other hand are designed to represent real world's entities like:

- 1) Ground and aerial vehicles (drones for example);
- 2) Humans;
- 3) Anti-aircraft systems.

**Definition 1. The Simulation Object (SO)**, as the core element of the simulation, consists of following components:

- 1) A logical state;
- 2) A physical state;
- 3) An identifier.

A logical state shall contain all data required to control behavior of a simulation object. For example, a logical state can be composed of information about whether the simulation object is still alive, or what is its destination.

A physical state shall contain all data required to visualize simulation object in simulation's world, example of such data might be its position and rotations about each of axes.

A SO's identifier shall be unique name of this object during simulation. Each pair of identifiers shall be comparable (on whether they are equal or not), and for the set of all identifiers the relation of order shall be established on.

The main reason for distinguishing logical state from physical state is to emphasize that they may differ even if they represent the same phenomenon. As an example let us consider a position of an object. In this case the physical state would represent actual values, one may say the values that are correct. The logical state in this case could represent data received from devices such as the GPS or some on-the-ground localization systems. It may happen that these two states will be very different from one another, specially if simulation object will not be able to establish connection with positioning system.

**Definition 2. The Passive Simulation Object (PSO)** is an extension of the simulation object. It can send messages,

but it cannot receive any. Passive simulation object is the simplest element that can take part in the simulation. Each PSO contains, despite what it has inherited from SO, following elements :

- 1) A physical state update function; (PSUF)
- 2) A physical state control function; It produces an input for the PSUF; (PSCF)
- 3) A messages generation function; (MGF)
- 4) A messages generation function trigger; (MGFT)
- 5) A (logical) state transition function (STF) and its trigger (STFT);

**Definition 3. The Active Simulation Object (ASO)** is an extension of the passive simulation object. Its capabilities extends PSO in a way that it is can receive messages. Each ASO contains, above what it has inherited from PSO, message receiving function.

#### C. Formal definitions of the basic model's components

Let  $I^{SO}$  denote a set of all (passive) simulation objects' identifiers. By  $I^{ASO} \subset I^{SO}$  we denote set of active simulation objects' identifiers, this is a subset of the set of all the identifiers. Sets  $LS_i, i \in I^{SO}$  and  $PS_i, i \in I^{SO}$  denotes accordingly a set of all possible logical states of simulation object, and a set of all physical states of the same object.

For all simulation objects there is a common set of basic physical state attributes  $PS_0$  defined as follows:

$$\forall i \in I^{SO} : PS_i = X_i^{PS} \times PS_0$$

$X_i^{PS}$  - a set of secondary physical state attributes used by the  $i^{th}$  simulation object.

Using the above definitions we can define following behavior controlling functions.

#### Definition 4. The Physical State Update Function

$PSUF_i : PS_i \times U_i \rightarrow PS_i, i \in I^{SO}$  - a physical state update function used by the  $i^{th}$  object.

$U_i, i \in I^{SO}$  - a set of vectors to control the  $i^{th}$  physical state.

#### Definition 5. The Physical State Control Function

$PSCF_i : PS_i \times LS_i \rightarrow U_i, i \in I^{SO}$  - a physical state control function used by the  $i^{th}$  object.

#### Definition 6. The State Transition Function Trigger

$STFT_{i,j} : LS_i \times PS_i \rightarrow \{true, false\}, j \in I_i^{STF}, i \in I^{SO}$  - a trigger of the  $j^{th}$  state transition function used by the  $i^{th}$  object.

#### Definition 7. The State Transition Function

$STF_{i,j} : PS_i \times LS_i \rightarrow LS_i, j \in I_i^{STF}, i \in I^{SO}$  - the  $j^{th}$  state transition function used by the  $i^{th}$  simulation object.

#### Definition 8. The Messages Generation Function Trigger

$MGFT_i : PS_i \times LS_i \rightarrow \{true, false\}$  - a messages generation function trigger used by the  $i^{th}$  simulation object.

In order to define last two functions we need to define one additional set:

$M^\infty$  - set of all messages.

#### Definition 9. The Messages Generation Function

$MGF_i : PS_i \times LS_i \times I^{SO} \rightarrow 2^{M^\infty}, i \in I^{SO}$  - a message generation function used by the  $i^{th}$  object.

**Definition 10. The Communication Capability Function**

$E_{comm} : \prod_{i \in I^{SO}} PS_i \rightarrow 2^{I^{SO} \times I^{SO}}$  – a communication capability function.

Interpretation:  $\forall i, j \in I^{SO} : \langle i, j \rangle \in 2^{I^{SO} \times I^{SO}} \iff i^{th}$  and  $j^{th}$  simulation objects can communicate with each other (they can exchange messages with one another).

**Definition 11. The mission objective of UAV swarm**

Let  $I_i^T \subset \mathbb{N}, i \in I^{SO}$  denote a set of tasks for the  $i^{th}$  object identifier.

Additionally let

$$M_i = \{T_{i,j} : PS_i \times LS_i \times \mathbb{Z} \rightarrow \{true, false\}\}_{j \in I_i^T}$$

denote a set of tasks for the  $i^{th}$  object. Interpretation of the function  $T_{i,j}$  is as follows. If physical and logical state, at the moment of function execution, meet the criteria of  $j^{th}$  task then the function returns *true*, otherwise it returns *false*. Having defined tasks for the  $i^{th}$  simulation object, a mission objective is the set of all tasks for every object, that is:

$$M = \bigcup_{i \in I^{SO}} M_i$$

A mission objective is considered as completed if all tasks for every object are accomplished.

$$\forall i \in I^{SO} \forall j \in I_i^T : T_{i,j}(ps_i(t), ls_i(t), t)$$

where:

$ps_i(t) : \mathbb{Z}_+ \rightarrow PS_i$  – a function returning physical state of the  $i^{th}$  object for a given simulation moment.

$ls_i : \mathbb{Z}_+ \rightarrow LS_i$  – a function returning logical state of the  $i^{th}$  object for a given simulation moment.

$t \in \mathbb{Z}_+$  – the simulation moment at which task condition is being checked.

**D. The simulation**

In the following section we will use defined herein notation for description of simulation steps. Notation:

$$\forall x \in X : func_1(x), i = x; X \subset \mathbb{Z}$$

is equal, in C# programming language, to:

```
foreach (int x in X)
{
    func1(x);
    i=x;
}
```

Let's define the basic swarm model simulation as follows. For subsequent  $t \in \mathbb{Z}_+$ , where  $\mathbb{Z}_+$  is the set of positive integer numbers:

- 1)  $\forall i \in I^{SO}$  update physical state of the  $i^{th}$  object:
 
$$u_i(t) = PSCF_i(ps_i(t), ls_i(t))$$

$$ps_i(t) = PSUF(ps_i(t), u_i(t))$$
- 2)  $\forall i \in I^{SO}$  check if the  $i^{th}$  object generates messages:
 
$$d_i(t) = MGFT_i(ps_i(t), ls_i(t))$$
- 3) Based on previous check, generate messages:
 
$$\forall i \in I^{SO} : d_i(t) == true \Rightarrow$$

$$M_i^\infty(t) = MGF_i(ps_i(t), ls_i(t), i)$$

- 4) Generate communication graph:

$$c(t) = E_{comm}(\prod_{i \in I^{SO}} ps_i(t))$$

- 5) Each **active** simulation object receives all messages from all objects it can communicate with:

$$\forall i \in I^{ASO} \forall \langle i, j \rangle \in c(t) \forall m \in M_j^\infty(t) : ls_i(t) = MRF_i(ps_i(t), ls_i(t), m)$$

- 6) Every simulation object updates its logical state:

$$\forall i \in I^{SO} \forall j \in I_i^{STF} : v_{i,j} = STFT_{i,j}(ls_i(t), ps_i(t))$$

$$\forall i \in I^{SO} \forall j \in I_i^{STF} \forall v_{i,j} == true : ls_i(t) = STF_{i,j}(ls_i(t), ps_i(t))$$

- 7) The logical and physical state of all objects at the end of each iteration becomes their initial state in the next iteration:

$$ls_i(t+1) = ls(t)$$

$$ps_i(t+1) = ps_i(t)$$

**V. THE EXTENDED MODEL WITH STIMULI**

The basic model was extended by adding sets of stimuli. The main reason for this was to allow the simulation objects to interact with fragments of the environment rather than directly interact with each other. All definitions from the basic model remains unchanged.

**A. The goal**

Main goals of the extended model were:

- 1) To allow an interference in communication between the objects to occur;
- 2) To add an influence of observer's position on content of received message. For example we can consider a loudness of a sound, that will be perceived differently by objects located in different places.

**B. The description of extended model's components**

All the definitions from IV-C stays unchanged, below are listed only new elements of the model.

**Definition 12. The stimulus**

A stimulus is a carrier of messages. By analogy we can exemplify it as radio waves.

**Definition 13. The Passive Simulation Object using Stimuli (PSOuS)**

A passive simulation object using stimuli is an extension of the PSO, by allowing it to emit stimuli based on previously generated messages.

It keeps limitations from the PSO, so it cannot receive any messages. Due to the above mentioned reason, it cannot receive any stimuli as well.

Each PSOuS consist of (regardless of what it has inherited from the PSO):

- 1) Stimulus emitting functions;
- 2) Selector of stimulus emitting function.

**Definition 14. The Active Simulation Object using Stimuli (ASOuS)**

A ASOuS is an extension of both the PSOuS and the ASO. It can generate messages, emit stimuli, receive them and receive messages.

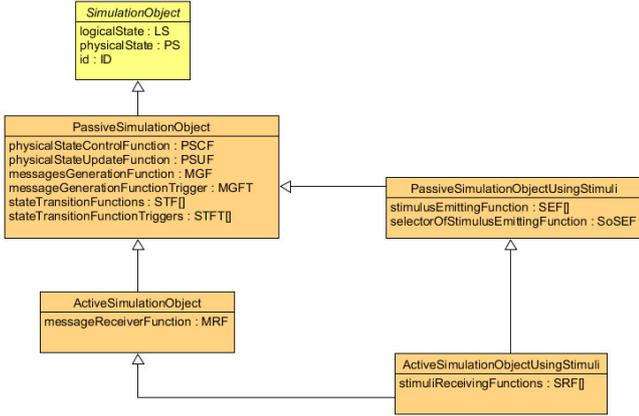


Fig. 1. Class diagram with the hierarchy of simulation object types.

Each ASOuS despite of what it has inherited from the PSOuS and the ASO, has stimuli receiving function.

A class diagram presenting the simulation object types hierarchy is presented on the figure 1.

### C. Formal definitions of the extended model's components

Let  $I^S$  denote a set of stimuli types identifiers, and by  $S_i, i \in I^S$  we denote a set of  $i^{th}$  type stimuli. Each set of stimuli must define following operations:

$+$ :  $S_i \times S_i \rightarrow S_i$  – the addition of the  $i^{th}$  type stimuli;  
 $P_i$ :  $PS_0 \times S_i \rightarrow S_i$  – the perception of the  $i^{th}$  type stimulus;  
 To clarify, each set of stimuli defines above operations only for its own type. Additionally, each set of stimuli must define a neutral element of itself.

$$\forall i \in I^S \exists ! e_i \in S_i \forall s \in S_i : +_i(s, e_i) = s$$

Finally, there must be a definition of perception operation on a subset of stimuli:

$$P_i^2 : PS_0 \times 2^{S_i} \rightarrow S_i, i \in I^S$$

### Definition 15. Stimulus Emitting Function

Let's denote a set of all stimulus emitting functions kept by  $i^{th}$  simulation object as  $SEM_i^\infty$ . Formal definition of the  $SEM_i^\infty$  set is as follows:

$\{SEM_{i,j,l} : PS_i \times LS_i \times M^\infty \rightarrow S_j\}_{l \in \mathbb{Z}_+, j \in I^S, i \in I^{SO}}$   
 $SEM_i^\infty$  – a set of all stimulus emitting functions kept by the  $i^{th}$  simulation object.

$SEM_{i,j,l}$  – a  $l^{th}$  stimulus emitting function being used by the  $i^{th}$  simulation object. The stimulus emitted by this function is of the  $j^{th}$  type.

### Definition 16. Selector of Stimulus Emitting Function

$SoSEM_i : PS_i \times LS_i \times M^\infty \times 2^{SEM_i^\infty} \rightarrow 2^{SEM_i^\infty}, i \in I^{SO}, j \in I^S$  – a selector of stimulus emitting function used by the  $i^{th}$  object.

### Definition 17. Stimuli receiving function

$SRF_i^\infty = \{SRF_{i,j,l} : PS_i \times LS_i \times 2^{S_j} \rightarrow M^\infty\}_{l \in \mathbb{Z}, i \in I^{SO}, j \in I^S}$

$SRF_i^\infty$  – a set of all stimuli receiving functions used by the

$i^{th}$  object.

$SRF_{i,j,l}$  – a  $l^{th}$  stimuli receiving function kept by the  $i^{th}$  object. This function accepts stimuli of the type  $j$ .

### Definition 18. Functional assigning corresponding stimuli type to emitters and receivers

$S^T : SRF_{i,j} \rightarrow I^S, i \in I^{SO}, j \in I^S$  – a functional assigning a stimulus type to a stimuli receiving function.  
 $S^T : SEM_{i,j} \rightarrow I^S, i \in I^{SO}, j \in I^S$  – a function assigning stimulus type to a stimulus emitting function.

### D. The simulation

All the definitions from IV-D remains unchanged. In order to describe the simulation process of the extended model, first we need to define the following function:

### Definition 19. The stimuli set at simulation moment function

$s_i : \mathbb{Z}_+ \rightarrow 2^{S_i}, i \in I^S$  – a function assigning each subsequent simulation moment its corresponding set of  $i^{th}$  type stimuli.

The simulation is performed according to the following steps.

For each subsequent  $t \in \mathbb{Z}_+$ :

- 1) See the step 1 of IV-D;
- 2) See the step 2 of IV-D;
- 3) See the step 3 of IV-D;
- 4) For every message generated by each simulation object select a set of emitters:  
 $\forall i \in I^{SO} \forall m \in M_i^\infty(t) : S_i^{SEL}(t) = S_i^{SEL}(t) \cup \{ \langle SoSEM_i(ls_i(t), ps_i(t), m, SEM_i^\infty), m \rangle \}$
- 5) Using each selection made in previous step generate messages:  
 $\forall i \in I^{SO} \forall \langle S^{SEM_s}, m \rangle \in S_i^{SEL}(t) \forall sem \in S^{SEM_s} : j = S^T(sem), s_j(t) = s_j(t) \cup \{ sem(ps_i(t), ls_i(t), m) \}$
- 6) With every stimuli receiver of each object perform perception operation on the appropriate set of stimuli:  
 $\forall j \in I^S \forall i \in I^{SO} \forall srf \in SRF_i^\infty : S^T(srf) == j \Rightarrow m_i^r(t) = m_i^r(t) \cup srf(ps_i(t), ls_i(t), s_j(t))$
- 7) Each simulation object receives all messages generated by its receivers in the previous step:  
 $\forall i \in I^{SO} \forall m \in m_i^r(t) : ls_i(t) = MRF_i(ps_i(t), ls_i(t), m)$
- 8) See the step 6 of IV-D;
- 9) See the step 7 of IV-D.

## VI. THE FRAMEWORK ARCHITECTURE

The framework architecture will be described according to [14]. The entire tool for the simulation of the UAV swarms can be divided into two web services and a client application. The client application is responsible for performing the simulation. Two web services provided by the framework, the first one to register every simulation moment, and the second one to visualize results of the simulation in 3D, are designed to reconstruct the course of the simulation later on (with vastly lower computational cost).

In the following architecture description we will primarily focus on the client application, because it is the only element that require implementation from a framework's user.

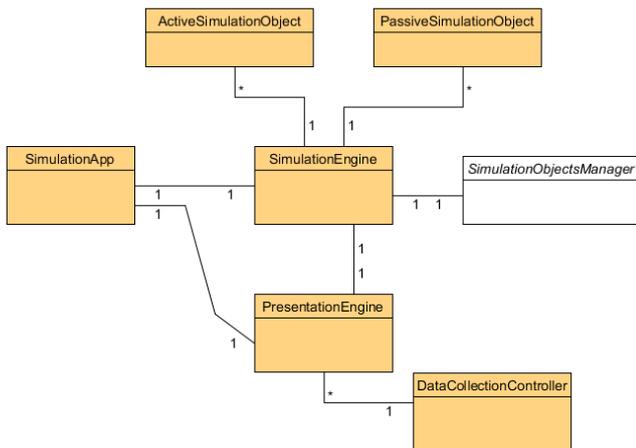


Fig. 2. The diagram of the main conceptual classes

### A. The logical view

The Figure 2 shows main conceptual classes that allows performing and saving simulation of the basic model results. The classes mentioned above are:

- **SimulationApp** – is the facade aimed to control the simulation. It is also responsible for management of the threads required to perform the simulation;
- **PresentationEngine** – is responsible for communication with the web service and saving simulation state on a server;
- **SimulationEngine** – supervises the course of the simulation and communicates with the presentation engine;
- **SimulationObjectsManager** – is the main element responsible for calculations needed by the simulation. It is also the element of which implementation quality impacts the performance of the simulation process the most;
- **DataCollectionController** – is the class responsible for registering the subsequent simulation moments;
- **ActiveSimulationObject** – is the class that represents the active simulation object. It contains all the elements depicted on the Figure 1
- **PassiveSimulationObject** – is the class that represents the passive simulation object. It contains all the elements depicted on the Figure 1.

The **ActiveSimulationObject** (usingStimuli) and **PassiveSimulationObject** (usingStimuli) classes are the only the components that requires further implementation.

### B. The process view

In the client module we can distinguish 2 main processes, as is depicted in Figure 3. The first process calculates subsequent states of the simulation, the second one communicates with the web services so the results from the first one can be saved.

It is worth mentioning that the above perspective is simplified, it doesn't involve any optimization (such as parallelization of calculation for every simulation object).

### C. The physical view

From the physical point of view, the client module instances can be located on the same device as the server is. Although, it is suggested to not share location of both modules.

Multiple client modules can be instantiated on one computer. It is also highly recommended for the computer that the database will be on to have sufficiently fast storage drive. It should be remembered that accessing the drive will occur much more often when reconstructing (visualizing) simulation, than when it is being performed (registered). The exemplary configuration is presented in the Figure 4.

### D. The developer view

The framework can be divided into 2 main modules, the client module that is doing the calculation, and the web module that registers the simulations, and it allows to access them later on.

The framework user needs to implement components such as functions to control physical state and messages generation/receiving function, they are all located in client module.

The client module consists of a set of dynamically loaded libraries (DLLs), while the web module is made of a WAR file that can be deployed on webservers like Payara[15] or Glassfish[16].

### E. The scenarios

The simplified scenario of the basic model's simulation was presented in the Figure 5. It is important to notice that steps 1-5 include, omitted on the picture, returning values needed for presentation layer from a manager.

Similarly the extended model's simulation scenario will only differ in number of steps it requires.

## VII. THE FRAMEWORK IMPLEMENTATION

The framework for the UAV swarms simulation was implemented in two technologies. The client module was implemented using C# (.NET Framework 4.7.2), and the web module was developed using JEE with frontend PrimeFaces framework in version 6.2. A 3D visualization of the simulation results is being performer with ThreeJS library.

The main reason why the client module was developed in the different language than the web module was the type erasure mechanism in the Java.

The following components are delivered as part of the framework:

- A basic implementation of physical state;
- A basic implementations of physical state control and update functions;
- A set of basic stimuli types;
- A tool for image-to-environment generation.

## VIII. AN EXEMPLARY SCENARIOS SIMULATION

The scenario of a UAV swarm mission consist of three elements:

- 1) Generating a map of a mission environment;
- 2) Configuring the simulation objects;

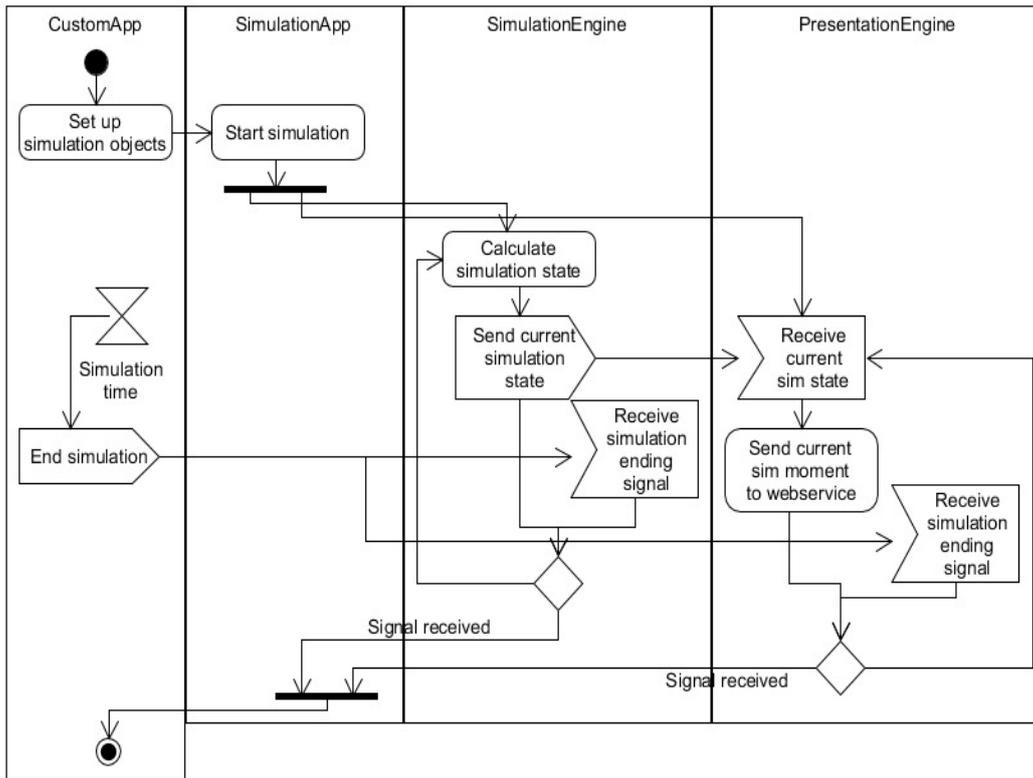


Fig. 3. The diagram of processes in the client module

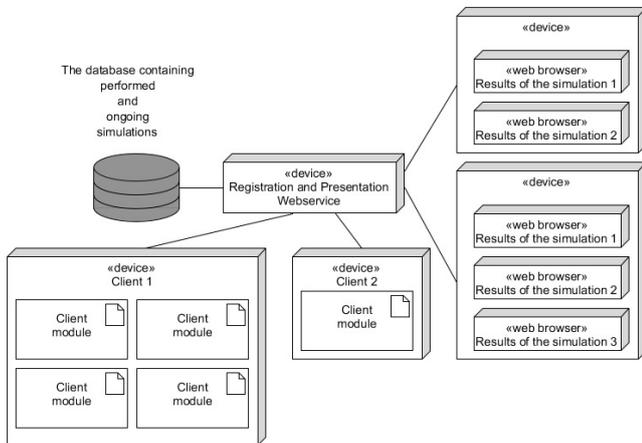


Fig. 4. The exemplary physical configuration

3) Setting up the mission objective for the simulation objects.

The mission environment map generation is a matter of two things. The first thing is to set parameters of the map, its width, depth, maximum heights etc. The second step is to select an image that represents the map. All these can be done with the provided tool (see Figure 7).

The simulation objects configuration is essentially about

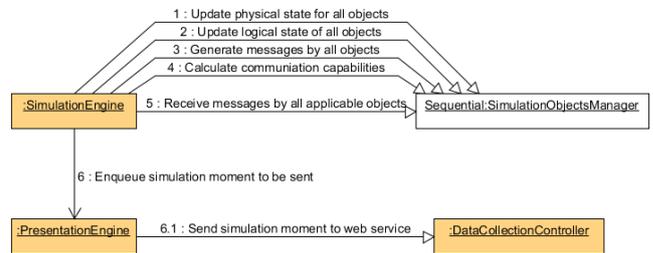


Fig. 5. The simulation of the basic model scenario

implementing functions described in IV or/and V.

The exemplary simulation scenario aimed to check whether a homogenous set of an unmanned aerial vehicles will be capable to reach a certain destination while avoiding environmental obstacles.

The environmental map was generated using bitmap depicted in the Figure 6.

Interpretation of the Figure 6 in the context of the map of environment is as follows:

- Height along the X axis will depend on the brightness of pixels along width of the picture;
- Height along the Z axis will depend on the brightness of pixels along height of the picture;

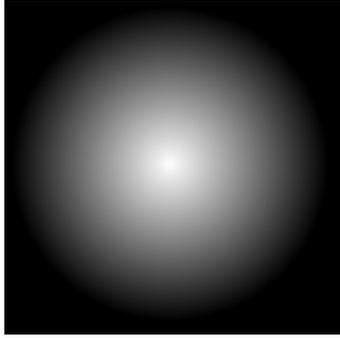


Fig. 6. The map of the mission environment used in the exemplary simulation scenario

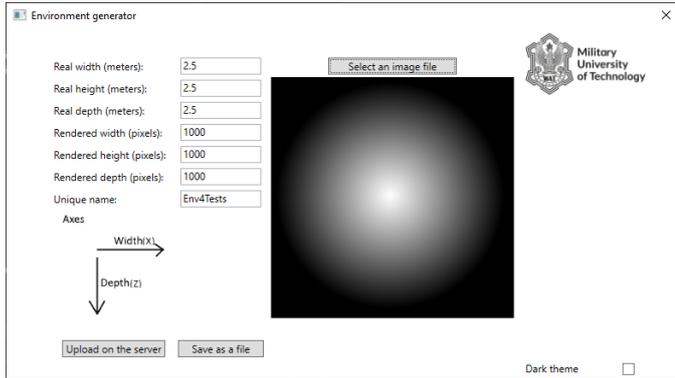


Fig. 7. The tool for mission environment generation

- The origin of the coordinate system (on the map) corresponds to the upper left corner in the picture.

The real world dimensions of the generated map were 2.5 meters by 2.5 meters. Maximum height (corresponding the white area in the Figure 6) was also set to 2.5 meters.

Four homogenous aerial vehicles have been placed on the map. Each of them had a goal to reach position (assuming xyz coordinate order) at point (2.5,0.1,2.3).

The simulation was performed using the extended model with one type of stimuli, let's call it collision type stimuli. A structure of the collision type stimulus is as follows:

$$S_1 = \{x : x \in R^3 \times \{true, false\}\}$$

The results confirmed possibility of performing the simulation with scenario described earlier. Three out of four objects reached their destination, the one that did not make it failed because of a flaw in an algorithm and not because of the framework. Below are listed figures showing particular simulation moments. The Figure 8 shows the beginning of the simulation, while Figure 9 shows middle of the simulation and the Figure 10 shows the simulation state at the end of the simulation process.

It is worth to mention that a map of the environment can be generated from any bitmap file. For example, more “advanced” map was generated using the file from Figure 11

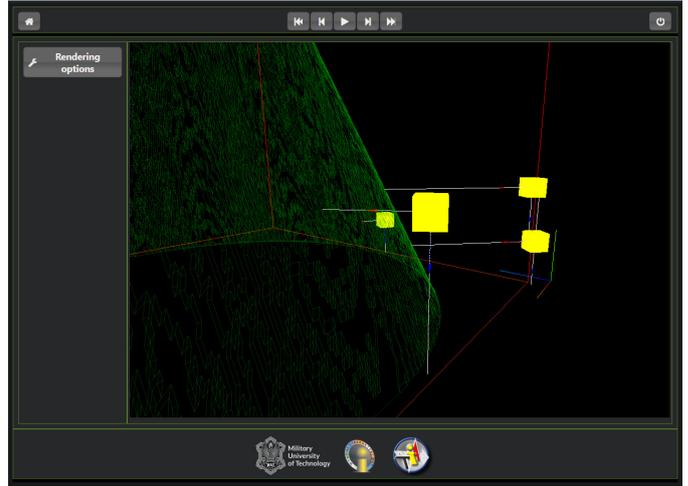


Fig. 8. The screenshot with results of a simulation in 14th simulation moment

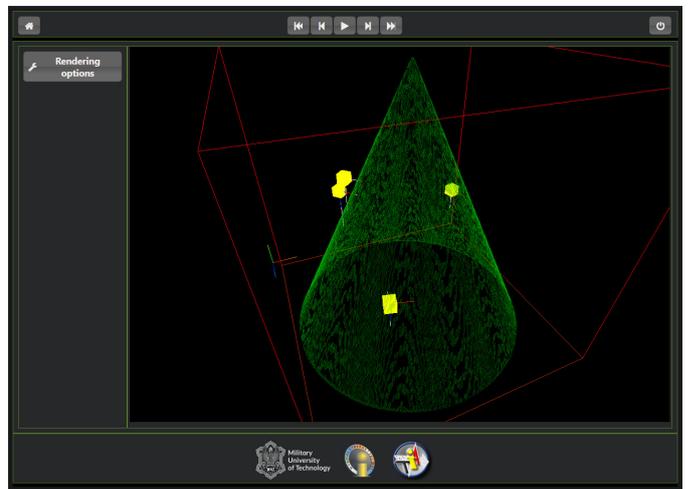


Fig. 9. The screenshot with results of a simulation in 109th simulation moment

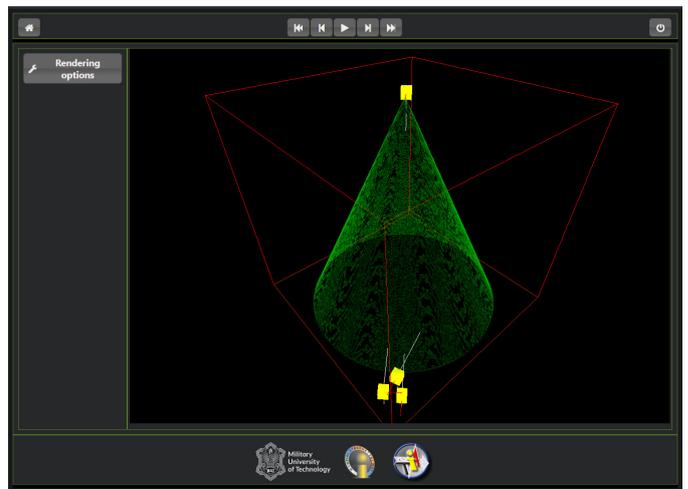


Fig. 10. The screenshot with results of a simulation in 431th simulation moment

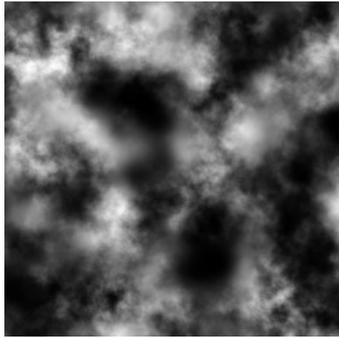


Fig. 11. An example of a mission area

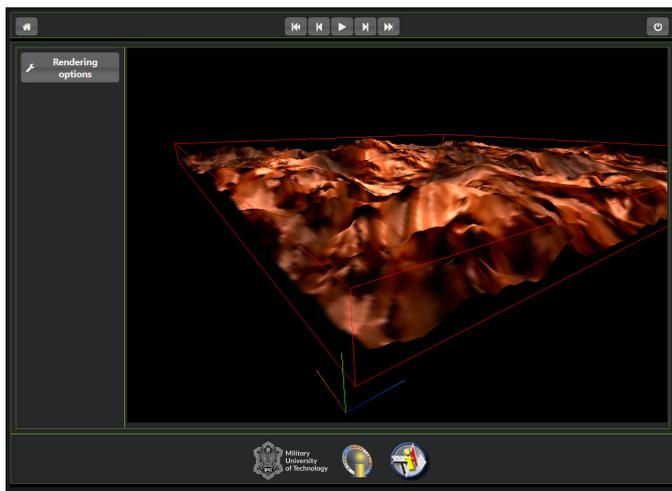


Fig. 12. A more advanced mission area – first view from the simulator

The mission environment generated from the Figure 11 was depicted on Figure 12 and Figure 13.

Each map can be rendered in two modes : solid and mesh. The map from Figure 12 shows the solid rendering mode, while Figure 13 depicts mesh rendering mode.

## IX. CONCLUSIONS

The goal of this work was to present the framework for the UAV swarm simulation with any scale and any type of tasks that may be set for it. Two simulation models, that are acceptable by the simulator, were presented, as well as their goals and limitations. The framework allows to perform simulation with user-defined map of environment, this is particularly useful when combined with the second model.

Results of an exemplary simulation scenarios were also presented.

There are two direction for future work on this framework. The first one is to design a script-like language so that a simulation objects configuration can be done without need of programmatic implementation. The second direction is to allow the simulation objects to be configured with languages other than C#, this would probably end up in designing a meta-framework for the tool presented here.

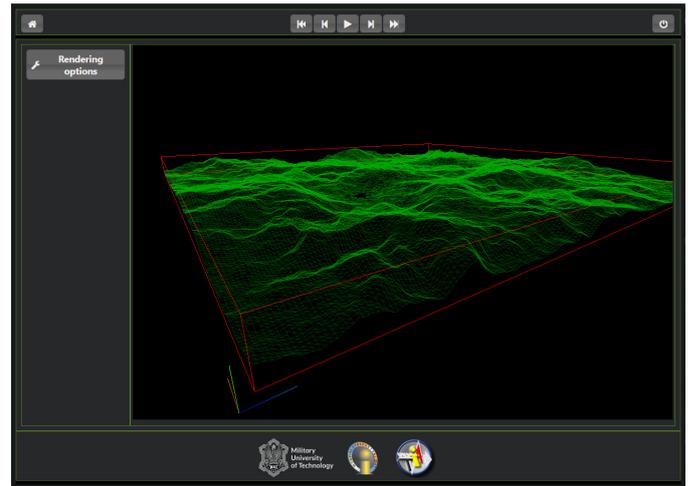


Fig. 13. A more advanced mission area – second view from the simulator

## REFERENCES

- [1] A. Kolling, P. M. Walker, N. Chakraborty, K. P. Sycara, and M. Lewis, "Human interaction with robot swarms: A survey," *IEEE Transactions on Human-Machine Systems*, vol. 46, pp. 9–26, 2016.
- [2] C. Gao, Z. Zhen, and H. Gong, "A self-organized search and attack algorithm for multiple unmanned aerial vehicles," *Aerospace Science and Technology*, 2016.
- [3] M. Kim, H. Baik, and S. Lee, "Response threshold model based uav search planning and task allocation," *Journal of Intelligent & Robotic Systems*, 2014.
- [4] H. Cheng, J. Page, and J. Olsen, "Dynamic mission control for uav swarm via task stimulus approach," *American Journal of Intelligent System*, vol. 2, pp. 177–183, 01 2013.
- [5] G. Wang, Q. Li, and L. Guo, "Multiple uavs routes planning based on particle swarm optimization algorithm," *International Symposium on Information Engineering and Electronic Commerce*, 07 2010.
- [6] W. Zhenhua, Z. Weiguo, S. Jingping, and H. Ying, "Uav route planning using multiobjective ant colony system," in *Conference on Cybernetics and Intelligent Systems*, pp. 797 – 800, 10 2008.
- [7] T. Ahmed, D. Feil-Seifer, T. Jiang, S. Jose, S. Liu, and S. Louis, "Development of a swarm uav simulator integrating realistic motion control models for disaster operations," 10 2017.
- [8] M. Brust, G. Danoy, P. Bouvry, D. Gashi, H. Pathak, and M. P. Goncalves, "Defending against intrusion of malicious uavs with networked uav defense swarms," pp. 103–111, 10 2017.
- [9] J. Wang, Y. Tang, J. Kavalen, A. Abdelzaher, and S. P. Pandit, "Autonomous uav swarm: Behavior generation and simulation," in *Conference on Unmanned Aircraft Systems (ICUAS)*, pp. 1–8, 06 2018.
- [10] S. Rasmussen, J. Mitchell, P. Chandler, C. Schumacher, and A. Smith, "Introduction to the multiuav2 simulation and its application to cooperative control research," pp. 4490 – 4501 vol. 7, 07 2005.
- [11] R. Garcia and L. Barnes, "Multi-uav simulator utilizing x-plane," *Journal of Intelligent and Robotic Systems*, vol. 57, pp. 393–406, 01 2010.
- [12] J. Bachrach, J. Mclurkin, and A. Grue, "Protoswarm: A language for programming multi-robot systems using the amorphous medium abstraction," vol. 2, pp. 1175–1178, 01 2008.
- [13] F. Bullo, J. Cortés, and S. Martínez, *Distributed Control of Robotic Networks*. Applied Mathematics Series, Princeton University Press, 2009. Electronically available at <http://coordinationbook.info>.
- [14] P. Kruchten, "Architectural blueprints – the 4+1 view model of software architecture," *IEEE Software* 12, 1995.
- [15] "<https://www.payara.fish/>"
- [16] "<https://javaee.github.io/glassfish/>"

# Using Relay Nodes in Wireless Sensor Networks: A Review

Mustapha Reda Senouci  
Ecole Militaire Polytechnique,  
BP 17, 16046, Bordj El-Bahri,  
Algiers, Algeria  
Email: mrsenouci@gmail.com

Mostefa Zafer  
Ecole nationale Supérieure d'Informatique,  
BP 68M, 16309, Oued-Smar,  
Alger, Algérie,  
Email: m\_zafer@esi.dz

Mohamed Aissani  
Ecole Militaire Polytechnique,  
BP 17, 16046, Bordj El-Bahri,  
Algiers, Algeria  
Email: maissani@gmail.com

**Abstract**—To extend the lifetime of wireless sensor networks, recent works suggest the use of relay nodes. This paper surveys and examines representative approaches dealing with relay nodes deployment. It also discusses their shortcomings and presents a comparative study. Additionally, this paper provides a set of remarks and recommendations to improve the usage of relay nodes in wireless sensor networks and highlights open issues that need further investigation.

## I. INTRODUCTION

A WIRELESS Sensor Network (WSN) is composed of Sensor Nodes (SNs) and Collector Nodes (CNs), deployed in a well-defined geographical area, called Region of Interest (RoI), to monitor the occurrence and/or evolution of a target event [1]. Each SN is responsible for collecting data associated with this event, via its sensing unit, and communicating them, using its wireless communication interface, to one of the CNs, directly, if the latter is in its communication range, or through a multi-hop routing, with the contribution of other intermediate SNs [1].

The successful completion of the control/monitoring mission assigned to the WSN requires that the network should be deployed and managed in a rigorous manner, which guarantees the RoI coverage and the WSN connectivity throughout the assigned mission [1]. The coverage must be of a predetermined order  $k$  ( $k \geq 1$ ), where each point of the RoI is covered by at least  $k$  SNs. Likewise, the connectivity must be of a predetermined degree  $l$  ( $l \geq 1$ ), which means that each SN has  $l$  disjoint paths, connecting it to the CNs. Usually, a redundancy in coverage and connectivity is provided ( $k, l \geq 2$ ) to ensure fault-tolerance.

In the WSN, the SNs are responsible, on the one hand, for collecting data related to the target event, and on the other hand, for forwarding it to the CNs. These two energy-hungry activities limit the lifetime of SNs, and lead sometimes, to the loss of coverage and/or connectivity [1]. Moreover, SNs have limited communication range, which requires, in most cases, the use of multi-hop routing, where packets go through multiple SNs before reaching the CNs, which increases the overall delivery latency and data loss rate.

To overcome the aforementioned challenges, some recent WSNs architectures [2], [3] include, in addition to SNs and CNs, relay nodes (RNs) that benefit from extended energy

autonomy and a communication range greater than that of SNs. In WSNs, these RNs are intended, in particular, to preserve [4], [5] or to restore [6], [7] the network connectivity, by actively participating in the forwarding of the collected data from the SNs to the CNs. This allows to balance the load between the SNs and the RNs [8], [9], [10], and thus prolong the network lifetime.

Nevertheless, the realistic and efficient utilization of RNs in WSNs is a long process that is in its early stage, and it currently faces several challenges [11], [12], [13]. It is a process consisting of several phases. In the first place, the appropriate usage mode of RNs should be selected. After that, the next phase chooses the topology of the network, according to the context, the constraints, and the desired objectives, before starting the RNs deployment phase. This latter is an NP-Hard problem [11], [12], [13] that consists of determining the number and the positions of RNs in the RoI. Finally, to extend the lifetime of the deployed WSN, a last phase designed to optimally manage the built topology should be considered.

In this paper, we clearly define the contours of the problem of using RNs in WSNs, by surveying and discussing representative approaches dealing with RNs deployment. This survey, intended to help researchers to quickly understand existing works, is completed by a set of remarks and recommendations to improve the usage of RNs in WSNs.

The rest of this paper is organized as follows. Section II presents the operating modes that describe the possible situations of RNs usage that can occur in practice. Then, the objectives and constraints of the problem at hand are identified in Section III. Section IV is devoted to the description of the possible WSN topologies in the presence of RNs. Section V surveys existing RNs deployment methods, and discusses their underlying assumptions. Next, Section VI summarizes the reviewed approaches, discusses their shortcomings, presents a comparative study, and highlights open issues. Finally, Section VII concludes the paper.

## II. OPERATING MODES

Guaranteeing the WSN connectivity throughout the planned monitoring/control mission remains the main motivation for RNs usage, despite their relatively high cost compared to that of SNs [14], [15], because the loss of connectivity may lead

to the failure of the mission assigned to the network. Ensuring network connectivity is achieved through two modes of RNs usage: reactive mode and proactive mode.

#### A. Reactive Mode

This mode of usage occurs when the WSN, consisting of SNs and CNs, is already deployed in the RoI (Fig. 1(a)). As time goes by, the battery-powered SNs will gradually be energy-exhausted and begin to disappear from the network, thereby leaving in their places what is commonly called coverage voids [1]. The number and sizes of these latter will expand gradually, causing sometimes the partitioning of the WSN into completely disjointed subnetworks (Fig. 1(b)). In this situation, some still operational SNs cannot communicate their data to any CN. It is at this moment that RNs should be deployed at specific locations [16], [17], [18], to restore the network connectivity, ensuring that each SN is able again to transmit its data to at least one CN (Fig. 1(c)). Thus, the deployed RNs act as gateways that interconnect the pieces of the network.

#### B. Proactive Mode

Sometimes, the use of the RNs along with SNs and CNs is considered at the network setup [16], [17], [18]. This is a precautionary measure by which the network administrator aims to preserve the coverage and the connectivity, throughout the planned monitoring/control mission [6], [7]. In fact, the RNs participate, from the beginning, in the routing of the data from SNs to CNs. In this way, SNs are, partially or totally, unloaded from the routing activity, which allows to extend their lifetimes, and thus, preserve the coverage quality of the RoI. On the other hand, the participation of RNs in data routing has a much lesser impact on their lifetimes, compared to SNs, because RNs have extended energy, which allows preserving the WSN connectivity.

The proactive usage of RNs is better than its reactive counterpart, in the sense that it maintains both coverage and connectivity, through load balancing between SNs and RNs starting from the network setup. However, it is not meant for emergency situations, where the restoration of connectivity should be immediate.

In practice, the choice between these two RNs usage modes is dictated by the order of priority given to each targeted objective (coverage, connectivity, cost, etc.), the constraints specific to the control mission (budget, duration, urgency, etc.), the constraints imposed sometimes by the RoI (a single area or several geographically distant areas) and by the functional characteristics of the employed SNs and RNs (communication range, storage capacity, etc.). The next section identifies the objectives and constraints commonly considered in the literature.

### III. OBJECTIVES AND CONSTRAINTS

Connectivity is not the only reason behind the usage of RNs in WSNs. Some existing works not only seek to preserve simple connectivity (1-connectivity) but attempt to provide

fault-tolerance, by deploying a sufficient number of RNs, so that each SN will have  $l$  ( $l \geq 2$ ) disjoint paths to forward its data to CNs [4], [5], [9], [10], [14], [19], [20], [21]. In this way, it would be possible to substitute, in the opportunistic moment, a broken or overloaded path, by another more appropriate path.

In addition to connectivity and fault-tolerance, the usage of RNs has other objectives such as **(1)** the minimization of cost, through the minimization of the number of RNs to be deployed [4], [5], [11]; **(2)** the minimization of the packet delivery delay [8], [13], [22], [23], [24], by ensuring that the routing paths are as short as possible; **(3)** the maximization of the network lifetime, notably by preserving the energy of SNs [2], [4], [11], [25] and/or by using RNs powered by green energy (where they harvest large amounts of ambient energy) [5], [16], [26], [27], and **(4)** the maximization of the communications links quality [22], [27], [28], which allows to avoid or to minimize the packet re-transmission operations that have a negative impact on the energy consumption and on the packets delivery delay [1].

Sometimes, some of the above-mentioned objectives are taken into account in the form of constraints. These latter could be related to the budget allocated to the mission, and are usually implemented by limiting the number of RNs to be deployed [2], [3], [17], [27]. Considered constraints could also be associated with the real-time nature of the intended application, and are carried out by requiring that the packets delivery delay is always below a tolerable threshold [8], [13], [29]. Other types of constraints are imposed by the RoI, such as limiting the possible positions of RNs [3], [4], [5], [8], [12], [19]. This last constraint often occurs in real-life applications, because the RoI can include hostile or inappropriate places, where it would be difficult, or even impossible, to place the RNs. Also, the positions to be occupied by the RNs are always dependent on the placement of SNs and CNs in the RoI, as the main role of the RNs is to restore and/or maintain the WSN connectivity.

Of course, some objectives are contradictory, such as fault-tolerance and deployment cost, and some others are perfectly correlated, such as the quality of communication links and the energy consumption of SNs and RNs. The definition of an order of priority among these objectives, which takes into account the various constraints and the correct relationship between them, remains necessary for efficient usage of RNs. The next section provides an analysis of the different WSNs topologies considered in the literature to meet the desired objectives.

### IV. WSN TOPOLOGIES

The objectives, the constraints, as well as the RNs usage mode adopted to satisfy them, all must be considered when selecting, thereafter, the adequate topology of the WSN. This topology, which defines the communications among the various nodes and the role of SNs and RNs in the data routing operation, have three possible forms, namely, the 1-tier topology, the 2-tier topology, and the hybrid topology.

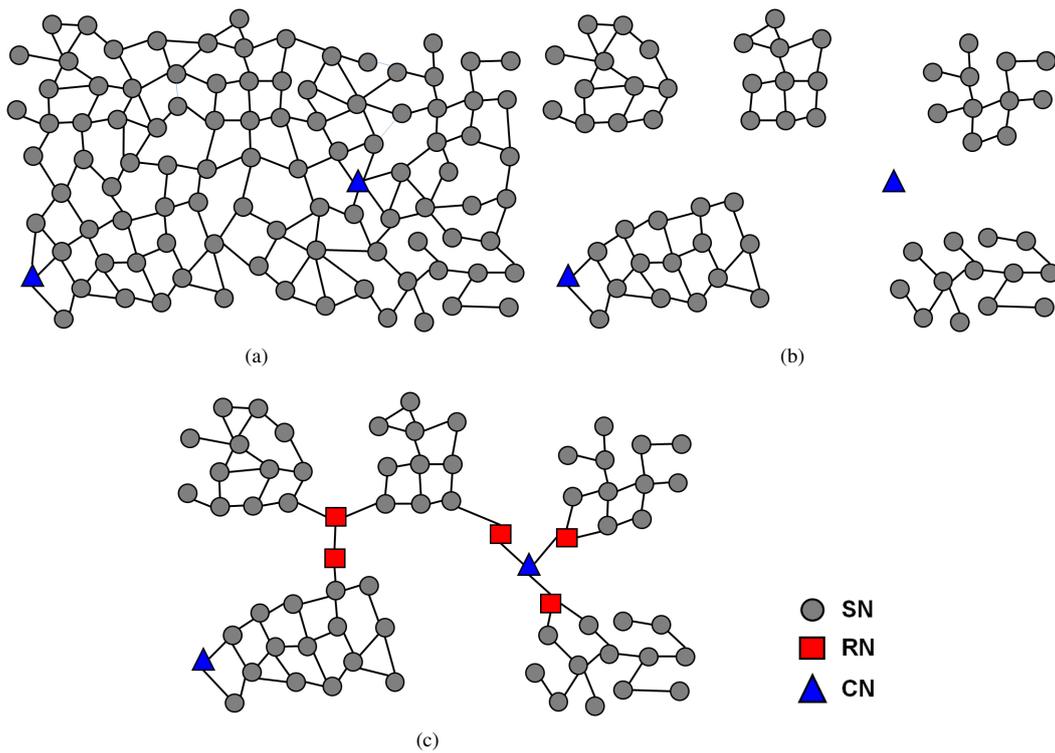


Fig. 1. Connectivity restoration by using RNs.

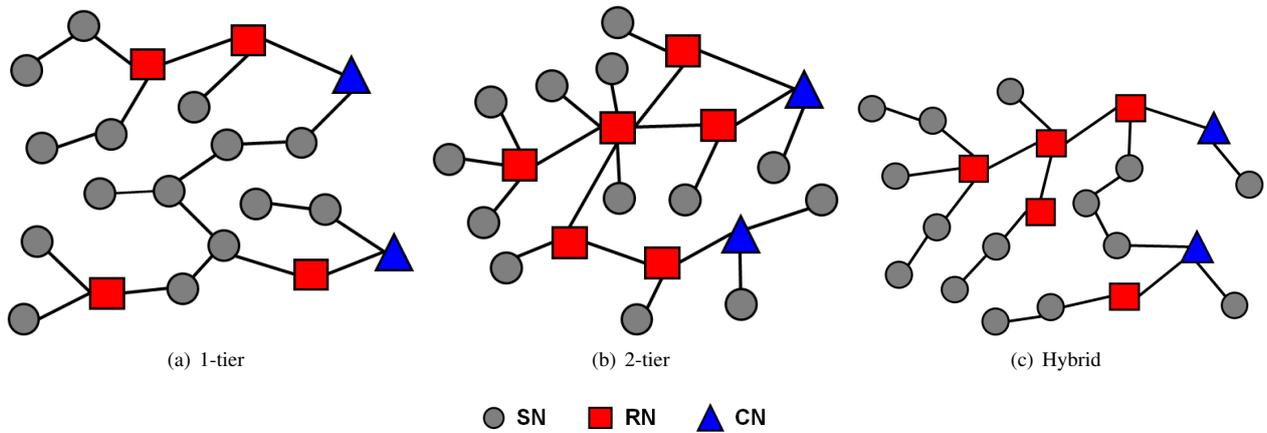


Fig. 2. WSN topologies composed of SNs, RNs, and CNs.

A. 1-tier topology

In this topology, shown in Fig. 2(a), SNs are involved, together with RNs, in the forwarding of data to CNs [6], [8], [11], [16], [17], [19], [28]. Thus, the intermediate nodes, forming the routing path connecting a SN to a CN, can be SNs and/or RNs. Typically, this 1-tier topology is selected when a reactive usage of RNs is assumed, more specifically, when RNs are merely used to restore the WSN connectivity. In this manner, RNs potential positions are limited and are located in the voids separating the pieces of the disconnected WSN [8],

[16], [18], [21].

B. 2-tier topology

In this topology, illustrated in Fig. 2(b), each SN deals only with the collection of data and transmits it to one of the CNs or, if necessary (no CN is within its communication range), to one of the RNs within its communication range. Subsequently, the RNs take care of routing this data to CNs [4], [5], [12], [13], [14]. Thus, the intermediate nodes, constituting the routing path connecting a SN to a CN, are necessarily RNs.

This 2-tier topology is seen as a clustering scheme of the WSN [4], [11], [14], [24], [30], where each cluster has a RN, acting as a cluster-head, and many SNs, acting as cluster members. This topology correlates much more with a proactive usage of RNs [5], [10], [27] because the positions to be occupied by the RNs should be well distributed over the entire RoI, so that each SN can communicate with at least one RN. In this way, it would be possible to build the necessary clusters based on RNs, such as all the cluster-heads are RNs [2], [4], [11], [14], [24], [30].

In comparison with the 1-tier topology, the 2-tier topology is more promising in terms of packets delivery delay and loss rate [8], [13], [22]. This is due to the fact that the routing paths, consisting only of RNs that have a communication range exceeding that of SNs, become shorter. Also, the 2-tier topology extends the lifespan of SNs [4], since these latter are totally unloaded from the routing task. In return, the 2-tier topology is costly, since its construction requires more RNs [4], [5], [19]. To address this gap, approaches adopting this 2-tier topology endeavor to minimize the number of used RNs.

### C. Hybrid topology

To take advantage of the aforementioned benefits of the 2-tier topology while minimizing the number of used RNs, some approaches [24], [25] adopt another topology, that we qualified it as *hybrid*. This hybrid topology resembles the 2-tier topology, except that the communication between a SN and its closest RN is performed, if necessary (RN is not in the communication range of the SN), via other intermediate SNs (see Fig. 2(c)). Thus, data collected by a SN can traverse a set of intermediate SNs, before reaching the first RN, which accomplish, together with some other RNs, the rest of the routing operation, similarly to the case when the 2-tier topology is used.

## V. RNS DEPLOYMENT APPROACHES

It should be noted that, unlike SNs that can be deployed randomly or deterministically [1], RNs are always deployed in a deterministic way, given the main following considerations: (i) RNs are responsible for forwarding data from SNs to CNs, whose positions must be computed according to the locations already occupied by SNs and CNs [31]; (ii) RNs are significantly more expensive than SNs. Thus, in comparison to a random deployment, a deterministic deployment allows to better optimize the total number of RNs.

The deterministic deployment of RNs consists in computing the number and the appropriate positions of the RNs in the RoI, which allow to reach the sought objectives and to respect the considered constraints while taking into account the assumed topology [4], [7], [28]. This problem has been shown to be computationally NP-hard [2], [3], [11], whose resolution involves two phases: a formulation phase and a resolution phase. The formulation phase describes, on the basis of the adopted assumptions, the relation between the different constraints and objectives, whereas the resolution phase, which is based on heuristics or meta-heuristics, selects from the

set of possible solutions, one solution that meets the desired objectives.

It should be noted that one of the most important assumptions that determine the soundness and practicability of a proposed deployment approach are those in relation with the adopted communication model. The latter describes, in a binary [2], [6], [11], [12], [17] or probabilistic manner [3], [16], [22], [26], [28], the quality of a communication link between two nodes, according to a set of parameters, which group in most of the proposed approaches, the distance between the communicating nodes [2], [3], [11], [12], [22], the transmission range (power) of nodes [2], [3], [11], [12], [22], the obstacles [32] as well as the medium of transmissions (radio, acoustic, etc.) [3].

In the related literature, there are different formulations of the deterministic deployment of RNs such as STP (Steiner Tree Problem) [4], [12], [17], [19], [22], CDS (Connected Dominating Set) [11], [25], SCP (Set Cover Problem) [8], [9], [13], [30], MST (Minimum Spanning Tree) [10] or ILP (Integer Linear Programming) [5], [7], [24], [26], [33]. Some other RNs deployment approaches [2], [14], [16], [28] come up with their proper formulation of the problem at hand. The formulated RNs deployment problem is solved through meta-heuristics such as the GA (Genetic Algorithms) [2], [20], the GSA (Gravitational Search Algorithm) [2], [18], [27], the DE (Differential Evolution) [27], the PSO (Particle Swarm Optimization) [18], the Column Generation (CG) algorithm [33] or other existing heuristics [34], [35], [36], [37], [38]. Some others RNs deployment approach [13], [14], [16], [26], [28] devise their proper heuristics to solve the problem at hand.

It should be remembered that the effectiveness of a RNs deployment approach depends on each of the above-mentioned phases, where the practicalness of an approach is strongly linked to the considered assumptions and constraints, as well as to the formulation phase describing their influence on the intended objectives. Moreover, the degree of satisfaction of the desired objectives, once they are correctly modeled in the formulation phase, depends on the efficiency of the adopted resolution method, which denotes its capacity to explore many possible solutions and to converge, quickly, towards a good-quality solution.

## VI. SYNTHESIS AND OPEN ISSUES

The introduction of RNs within WSNs in order to preserve or restore connectivity, in addition to other objectives, is a delicate task, which requires in the first place a clear definition of the objectives and an inventory of the various constraints, to determine the proper operating mode (reactive or proactive). Taking into account the latter, and without disregarding the constraints and the searched objectives, the appropriate topology could be selected. It is at this point that the real work of deployment begins, which consists of determining the number and positions of RNs, allowing the network to operate according to the chosen topology while respecting the objectives and constraints previously identified. This process

is illustrated in Fig. 3, which highlights the sequence and dependence among these phases.

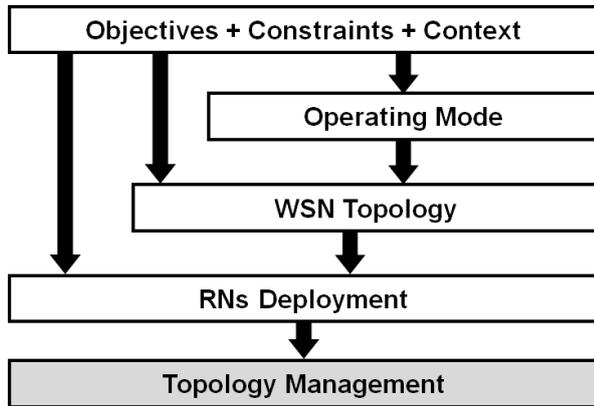


Fig. 3. Process of using RNs in WSNs.

Table I shows a comparative analysis of the most recent RNs deployment approaches. Although they start with well-defined objectives, some of these approaches have motivated the selection of the appropriate topology (1-tier, 2-tier, or hybrid) based only on the objectives. They overlooked the fact that this choice is tributary also to external constraints related to the envisaged application (budget and delay of deployment, allowed positions for RNs, urgency, etc.), which require to identify, in the first place, the appropriate operating mode (reactive or proactive) before thinking about the topology to be adopted. Due to the omission of this intermediate phase, the context of these approaches usage remains unclear.

In addition, these approaches suffer from a problem of practicalness, due to the omission of certain constraints having a quasi-permanent presence in real-life applications, or because of the adoption of unrealistic assumptions. For instance, excepting some works [3], [39], [30] that considered 3D or very specific RoI, all mentioned approaches are designed and/or evaluated under the assumption that the RoI is purely 2D, which makes them irrelevant for realistic 3D RoI, and limits their usage to very limited cases.

Furthermore, besides few works [32] that take into account, in a superficial way, the impact of obstacles, the wireless communications are modeled, in the others approaches, by very simple models that consider only the distance and the communication range of the nodes, omitting other important factors, such as obstacles, which can prevent the communications, or in the best cases, degrade their qualities. Also, some of these models are binary, forgetting the probabilistic nature of wireless communications, confirmed in several previous works.

Additionally, the constraint of RNs positions, which has a real and quasi-permanent presence, has been overlooked, or in the best cases, treated in a very superficial way, where all the allowed positions of RNs are chosen arbitrarily. However, in reality, the determination of these positions requires an analysis of the RoI. The impact of this gap will be more

significant in the case of purely 3D RoI, where the topography is the first factor that imposes this kind of constraints.

Finally, a serious problem should be mentioned and it concerns the basic idea behind the use of RNs in WSNs. Indeed, the intensive participation of RNs in data routing can exhaust, in a very fast manner, their limited residual energy, especially when the 2-tier topology is considered. This aspect has not been well dealt with and remains an open issue. To handle this situation, some approaches assume that a RN has an unlimited energy resource, which remains an unrealistic assumption (at least for now). Some other approaches have aimed to ensure  $k$ -connectivity to face the failure of RNs, but we believe that they need to be strengthened. More precisely, we believe that the process of exploiting RNs in WSNs should be fortified by a network topology management phase (see Fig. 3), designed in particular to optimize the RNs energy consumption, by using a load balancing strategy. To the best of our knowledge, no such complete solution has been proposed. Consequently, efficient topology management of a heterogeneous WSN, composed of SNs and RNs, remains an open issue.

## VII. CONCLUSION

Recent works suggest enhancing traditional WSNs architecture, consisting of SNs and CNs, by introducing RNs to provide reliable data transport from SNs to CNs. In this paper, we have presented and discussed the steps that constitute the process of exploiting RNs in WSNs. This process starts with the definition of the desired objectives and ends with the deployment phase, which consists in determining the number and the positions of RNs in the RoI. In the end, we have carried out a comparative study among the most recent RNs deployment strategies and pointed out the main shortcomings of existing works.

The identified shortcomings are related in particular to the practicalness of these works since most of them assumed that the RoI is 2D and adopted unrealistic communications models, omitting the undeniable impact of some factors on wireless communications, notably obstacles. Nevertheless, the most important remark is the absence of a topology management phase that should follow the deployment of the WSN, and which helps to optimize the resources of the WSN, especially the energy of RNs that are actively involved in data routing. This phase, which must further reinforce the measures taken by some recent works that have targeted fault-tolerance (by ensuring the  $k$ -connectivity of the network), should notably manage the load balancing among RNs. We believe that this phase is highly important and deserves immediate attention.

## REFERENCES

- [1] M. R. Senouci and A. Mellouk, *Deploying Wireless Sensor Networks: Theory and Practice*. Elsevier, 2016.
- [2] J. M. Lanza-Gutierrez and J. A. Gomez-Pulido, "A gravitational search algorithm for solving the relay node placement problem in wireless sensor networks," *International Journal of Communication Systems*, vol. 30, no. 2, 2015. doi: 10.1002/dac.2957

TABLE I  
A COMPARISON BETWEEN THE MOST RECENT RNS DEPLOYMENT APPROACHES.

Ref.	RoI	Operating Mode	Topology	Connectivity degree ( $k$ )	Objectives					Constraints			Deployment		
					RNs Number	RNs Energy	SNs Energy	Comm. Quality	Delivery delay	RNs positions	Delivery delay	RNs Numbers	Comm. Model	Form. Method	Resol. Method
[2]	2D	Proactive	1-tier	$k = 1$	✓					✓		Binary	New	GA + GSA	
[3]	Specific	Proactive	2-tier	$k = 1$	✓	✓	✓		✓			Probabilistic	ILP	New	
[4]	2D	Proactive	2-tier	$k \in \{1, 2\}$	✓	✓			✓			Binary	STP	[34]	
[5]	2D	Proactive	2-tier	$k \in \{1, 2\}$	✓	✓			✓			Binary	ILP	[36]	
[6], [15]	2D	Reactive	1-tier, 2-tier	$k = 1$	✓							Binary	STP	[35] + New	
[7]	2D	Reactive	1-tier	$k = 1$	✓							Binary	ILP	New	
[8], [13]	2D	Proactive	1-tier	$k = 1$	✓				✓	✓	✓	Binary	SCP	[38] + New	
[9], [40]	2D	Proactive	2-tier	$k \in \{1, 2\}$	✓							Binary	SCP	New	
[10]	2D	Proactive	2-tier	$k = 2$	✓							Binary	MST	New	
[11], [25]	2D	Proactive	1-tier + Hybrid	$k = 1$	✓	✓						Binary	CDS + ILP	[36], [37]	
[12]	2D	Proactive	2-tier	$k = 1$	✓				✓			Binary	STP + ILP	[36]	
[14]	2D	Proactive	2-tier	$k \in \{2, 3, \dots\}$	✓							Binary	New	New	
[16]	2D	Reactive	1-tier	$k = 1$	✓	✓						Probabilistic	New	New	
[41]	2D	Proactive	2-tier	$k = 1$		✓	✓	✓				Probabilistic	New	New	
[17], [18], [42]	2D	Reactive	1-tier	$k = 1$	✓					✓		Binary	STP	PSO + New	
[19]	2D	Proactive	1-tier	$k \in \{1, 2\}$	✓				✓			Binary	STP	[34]	
[20]	2D	Proactive	2-tier	$k \in \{2, 3, \dots\}$	✓				✓			Binary	ILP	GA	
[21]	2D	Proactive	1-tier, 2-tier	$k = 2$	✓							Binary	STP	[15]	
[22]	2D	Proactive	2-tier	$k = 1$	✓		✓	✓	✓			Probabilistic	STP	[34]	
[23]	2D	Proactive	2-tier	$k = 1$	✓	✓	✓		✓			Probabilistic	New	New	
[24]	2D	Proactive	Hybrid	$k = 1$	✓				✓			Binary	ILP	New	
[26]	2D	Proactive	2-tier	$k = 1$	✓	✓			✓			Probabilistic	ILP	New	
[27]	2D	Proactive	2-tier	$k = 1$	✓	✓	✓			✓		Probabilistic	ILP	DE + GSA	
[28]	2D	Proactive	1-tier	$k = 1$			✓			✓		Probabilistic	New	New	
[29], [43]	2D	Proactive	1-tier	$k = 1$	✓				✓	✓	✓	Probabilistic	STP + New	New	
[30]	Specific	Proactive	2-tier	$k = 1$	✓		✓					Probabilistic	SCP	New	
[32]	2D	Proactive	2-tier	$k = 1$	✓				✓			Probabilistic	ILP	New	
[33]	2D	Proactive	1-tier	$k = 1$	✓				✓	✓		Binary	ILP	CG	
[44]	2D	Proactive	1-tier	$k = 1$		✓	✓	✓				Probabilistic	ILP	New	
[39]	3D	Proactive	2-tier	$k = 1$	✓		✓		✓			Probabilistic	STP	[34]	
[45]	3D	Proactive	2-tier	$k \in \{2, 3, \dots\}$	✓		✓		✓			Probabilistic	STP	[34]	
[46]	2D	Proactive	2-tier	$k = 1$	✓	✓	✓					Probabilistic	New	New	

- [3] D. Wu, D. Chatzigeorgiou, K. Youcef-Toumi, S. Mekid, and R. Ben-Mansour, "Channel-Aware Relay Node Placement in Wireless Sensor Networks for Pipeline Inspection," *Transactions on Wireless Communications*, vol. 13, no. 7, pp. 3510–3523, 2014. doi: 10.1109/TWC.2014.2314120
- [4] D. Yang, S. Misra, X. Fang, G. Xue, and J. Zhang, "Two-Tiered Constrained Relay Node Placement in Wireless Sensor Networks: Computational Complexity and Efficient Approximations," *Transactions on Mobile Computing*, vol. 11, no. 8, pp. 1–13, 2012. doi: 10.1109/TMC.2011.126
- [5] S. Misra, N. E. Majd, and H. Huang, "Approximation Algorithms for Constrained Relay Node Placement in Energy Harvesting Wireless Sensor Networks," *Transactions on Computers*, vol. 63, no. 12, pp. 2933–2947, 2013. doi: 10.1109/TC.2013.171
- [6] H. Zeng and Z. Kang, "Relay Node Placement to Restore Connectivity in Wireless Sensor Networks," in *9th International Conference on Communication Software and Networks (ICCSN)*, Guangzhou, China, Dec. 2017. doi: 10.1109/ICCSN.2017.8230124 pp. 301–305.
- [7] G. Xiong, L. Hong, and Y. Guangyou, "Improving Energy Efficiency by Optimizing Relay Nodes Deployment in Wireless Sensor Networks," in *9th International Conference on Communication Software and Networks (ICCSN)*, Guangzhou, China, Dec. 2017. doi: 10.1109/ICCSN.2017.8230125 pp. 306–310.
- [8] C. Ma, W. Liang, and M. Zheng, "Set-Covering-based Algorithm for Delay Constrained Relay Node Placement in Wireless Sensor Networks," in *International Conference on Communications (ICC)*, Kuala Lumpur, Malaysia, July 2016. doi: 10.1109/ICC.2016.7510976 pp. 1–6.
- [9] C. Ma, W. Liang, M. Zheng, and H. Sharif, "A Connectivity-Aware Approximation Algorithm for Relay Node Placement in Wireless Sensor Networks," *Sensors*, vol. 16, no. 2, pp. 515–528, 2015. doi: 10.1109/JSEN.2015.2456931
- [10] Q. Chen, Y. Hu, Z. Chen, V. Grout, D. Zhang, H. Wang, and H. Xing, "Improved Relay Node Placement Algorithm for Wireless Sensor Networks Application in Wind Farm," in *International Conference on Smart Energy Grid Engineering (SEGE)*, Oshawa, Canada, Jan. 2013. doi: 10.1109/SEGE.2013.6707901 pp. 1–6.
- [11] D. Djenouri and M. Bagaa, "Energy Harvesting Aware Relay Node Addition for Power-Efficient Coverage in Wireless Sensor Networks," in *International Conference on Communications (ICC)*. London, UK: IEEE, Sep. 2015. doi: 10.1109/ICC.2015.7248303 pp. 86–91.
- [12] A. Chelli, M. Bagaa, D. Djenouri, I. Balasingham, and T. Taleb, "One Step Approach for Two-Tiered Constrained Relay Node Placement in Wireless Sensor Networks," *Wireless Communications Letters*, vol. 5, no. 4, pp. 448–451, 2016. doi: 10.1109/LWC.2016.2583426
- [13] C. Ma, W. Liang, and M. Zheng, "Delay Constrained Relay Node Placement in Two-tiered Wireless Sensor Networks: A Set-Covering-based Algorithm," *Journal of Network and Computer Applications*, vol. 93, pp. 76–90, 2017. doi: 10.1016/j.jnca.2017.05.004
- [14] K. Nitesh and P. K. Jana, "Relay Node Placement with Assured Coverage and Connectivity: A Jarvis March Approach," *Wireless Personal Communications*, vol. 98, no. 1, pp. 1361–1381, 2017. doi: 10.1007/s11277-017-4922-8
- [15] E. L. Lloyd and G. Xue, "Relay Node Placement in Wireless Sensor Networks," *Transactions on Computers*, vol. 56, no. 1, pp. 134–138, 2007. doi: 10.1007/s11276-006-0724-8
- [16] S. Xu, L. Jiang, C. He, and Q. Xi, "Relay Node Placement in Partitioned Wireless Sensor Networks with Guaranteed Lifetime," in *Global Communications Conference (GLOBECOM)*, Atlanta, USA, June 2013. doi: 10.1109/GLOCOM.2013.6831078 pp. 243–248.
- [17] C. Zhou, A. Mazumder, A. Das, K. Basu, N. Matin-Moghaddam, S. Mehrani, and A. Sen, "Relay Node Placement Under Budget Constraint," in *19th International Conference on Distributed Computing and Networking*, Varanasi, India, Jan. 2018. doi: 10.1145/3154273.3154302 pp. 1–6.
- [18] Y.-H. Xu, W.-G. Jiao, YinWu, and J. Song, "Variable-dimension swarm meta-heuristic for the optimal placement of relay nodes in wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 13, no. 3, pp. 1–15, 2017. doi: 10.1177/1550147717700895
- [19] S. Misra, S. D. Hong, G. L. Xue, and J. Tang, "Constrained Relay Node Placement in Wireless Sensor Networks: Formulation and Approximations," *IEEE/ACM Transactions on Networking*, vol. 18, no. 2, pp. 434–447, 2010. doi: 10.1109/TNET.2009.2033273
- [20] M. Azharuddin and P. K. Jana, "A GA-based approach for fault tolerant relay node placement in wireless sensor networks," in *Third International Conference on Computer, Communication, Control and Information Technology (C3IT)*, Hooghly, India, March 2015. doi: 10.1109/C3IT.2015.7060111 pp. 1–6.
- [21] W. Zhang, G. Xue, and S. Misra, "Fault-Tolerant Relay Node Placement in Wireless Sensor Networks: Problems and Algorithms," in *26th International Conference on Computer Communications*, Barcelona, Spain, May 2007. doi: 10.1109/INFCOM.2007.193 pp. 1649–1657.
- [22] M. Bagaa, A. Chelli, D. Djenouri, T. Taleb, I. Balasingham, and K. Kansanen, "Optimal Placement of Relay Nodes Over Limited Positions in Wireless Sensor Networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 4, pp. 2205–2219, 2017. doi: 10.1109/TWC.2017.2658598
- [23] W. Zhu, S. Xianhe, L. Cuicui, and C. Jianhui, "Relay Node Placement Algorithm Based on Grid in Wireless Sensor Network," in *Third International Conference on Instrumentation, Measurement, Computer, Communication and Control*, Shenyang, China, June 2013. doi: 10.1109/IMCCC.2013.65 pp. 278–283.
- [24] M. Shokrnezhad, V. Zolfaghari, and S. Khorsandi, "Relay Node Placement in Mission Critical Smart Grid Networks," in *7th International Symposium on Telecommunications*, Tehran, Iran, Sept. 2014. doi: 10.1109/ISTEL.2014.7000794 pp. 707–711.
- [25] D. Djenouri and M. Bagaa, "Energy-Aware Constrained Relay Node Deployment for Sustainable Wireless Sensor Networks," *Transactions on Sustainable Computing*, vol. 2, no. 1, pp. 30–42, 2017. doi: 10.1109/TSUSC.2017.2666844
- [26] Z. Zheng, L. X. Cai, R. Zhang, and X. S. Shen, "RNP-SA: Joint Relay Placement and Sub-Carrier Allocation in Wireless Communication Networks with Sustainable Energy," *Transactions on Wireless Communications*, vol. 11, no. 10, pp. 3818–3828, 2012. doi: 10.1109/TWC.2012.090312.120461
- [27] B. O. Ayinde and H. A. Hashim, "Energy-Efficient Deployment of Relay Nodes in Wireless Sensor Networks Using Evolutionary Techniques," *International Journal of Wireless Information Networks*, vol. 25, no. 2, pp. 157–172, 2018. doi: 10.1007/s10776-018-0388-1
- [28] M. Nikolov and Z. J. Haas, "Relay Placement in Wireless Networks: Minimizing Communication Cost," *Transactions on Wireless Communications*, vol. 15, no. 5, pp. 3587–3602, 2016. doi: 10.1109/TWC.2016.2523984
- [29] A. Bhattacharya and A. Kumar, "Delay Constrained Optimal Relay Placement for Planned Wireless Sensor Networks," in *18th International Workshop on Quality of Service (IWQoS)*, Beijing, China, Aug. 2010. doi: 10.1109/IWQoS.2010.5542760 pp. 1–9.
- [30] R. Liu, I. J. Wassell, and K. Soga, "Relay Node Placement for Wireless Sensor Networks Deployed in Tunnels," in *6th International Conference on Wireless and Mobile Computing, Networking and Communications*, Niagara Falls, Canada, Nov. 2010. doi: 10.1109/WIMOB.2010.5644984 pp. 144–150.
- [31] M. Zafer, M. R. Senouci, and M. Aissani, "Terrain Partitioning Based Approach for Realistic Deployment of Wireless Sensor Networks," in *International Conference on Computational Intelligence and Its Applications*, May 2018. doi: 10.1007/978-3-319-89743-1\_37 pp. 423–435.
- [32] F. M. Al-Turjman, A. E. Al-Fagih, W. M. Alsalih, and H. S. Hasanein, "A delay-tolerant Framework for Integrated RSNs in IoT," *Computer Communications*, vol. 36, no. 9, pp. 998–1010, 2013. doi: 10.1016/j.comcom.2012.07.001
- [33] A. Nigam and Y. K. Agarwal, "Optimal Relay Node Placement in Delay Constrained Wireless Sensor Network Design," *European Journal of Operational Research*, vol. 233, no. 1, pp. 220–233, 2014. doi: 10.1016/j.ejor.2013.08.031
- [34] L. T. Kou, G. Markowsky, and L. Berman, "A Fast Algorithm for Steiner Trees," *Acta Informatica*, vol. 15, no. 2, pp. 141–145, 1981. doi: 10.1007/BF00288961
- [35] Z.-X. Yang, X.-Y. Jia, J.-Y. Hao, and Y.-P. Gao, "Geometry experiment algorithm for steiner minimal tree problem," *Journal of Applied Mathematics*, vol. 2013, pp. 507–508, 2013. doi: 10.1155/2013/367107
- [36] "Ilog cplex: Software for mathematical programming and optimization," <http://www.ilog.com/products/cplex/>, 2002.
- [37] D. Kim, Z. Zhang, X. Li, W. Wang, W. Wu, and D.-Z. Du, "A Better Approximation Algorithm for Computing Connected Dominating Sets in Unit Ball Graphs," *Transactions on Mobile Computing*, vol. 9, no. 8, pp. 1108–1118, 2010. doi: 10.1109/TMC.2010.55
- [38] Thomas H. Cormen, C. E. Leiserson, C. L. Rivest, and C. Stein, *Introduction to Algorithm*. MIT Press and McGraw-Hill, 2009.

- [39] M. Zafer, M. R. Senouci, and M. Aissani, "A Practical Data Driven Approach for the Deployment of WSNs on Realistic Terrains," *Transactions on Emerging Telecommunications Technologies*, Jan. 2019. doi: 10.1002/ett.3558
- [40] C. Ma, W. Liang, M. Zheng, and H. Sharif, "A Novel Local Search Approximation Algorithm for Relay Node Placement in Wireless Sensor Networks," in *Wireless Communications and Networking Conference (WCNC)*, New Orleans, USA, June 2015. doi: 10.1109/WCNC.2015.7127693 pp. 1536–1541.
- [41] F. Al-Turjman, "Optimized Hexagon-based Deployment for Large-Scale Ubiquitous Sensor Networks," *Journal of Network and Systems Management*, vol. 26, no. 2, pp. 255–283, 2017. doi: 10.1007/s10922-017-9415-2
- [42] C. Zhou, A. Mazumder, A. Das, K. Basu, N. Matin-Moghaddam, S. Mehrani, and A. Sen, "Relay Node Placement Under Budget Constraint," *Pervasive and Mobile Computing*, vol. 53, pp. 1–12, 2019. doi: 10.1016/j.pmcj.2018.12.001
- [43] A. Bhattacharya and A. Kumar, "A Shortest Path Tree Based Algorithm for Relay Placement in a Wireless Sensor Network and Its Performance Analysis," *Computer Networks*, vol. 71, pp. 48–62, 2014. doi: 10.1016/j.comnet.2014.06.011
- [44] F. Al-Turjman, "QoS-aware Data Delivery Framework for Safety-inspired Multimedia in Integrated Vehicular-IoT," *Computer Communications*, vol. 121, pp. 33–43, May 2018. doi: 10.1016/j.comcom.2018.02.012
- [45] M. Zafer, M. R. Senouci, and M. Aissani, "Fault-Tolerant Data Transport Backbone for 3D Wireless Sensor Networks," *Transactions on Emerging Telecommunications Technologies*, May 2019. doi: 10.1002/ett.3660
- [46] F. Al-Turjman, "Cognitive Routing Protocol for Disaster-inspired Internet of Things," *Future Generation Computer Systems*, vol. 42, pp. 317–334, March 2017. doi: 10.1016/j.future.2017.03.014

# Inference of driver behavior using correlated IoT data from the vehicle telemetry and the driver mobile phone

Daniel Alves da Silva, José Alberto Sousa Torres, Alexandre Pinheiro, Francisco L. de Caldas Filho, Fabio L. L. Mendonça, Bruno J. G. Praciano, Guilherme de Oliveira Kfourir and Rafael T. de Sousa Jr.

*Department of Electrical Engineering  
University of Brasília, Brasília, Brasil*

{daniel.alves, albero.torres, alexandre.pinheiro, francisco.lopes,  
fabio.mendonca, bruno.praciano, guilherme.kfourir}@redes.unb.br, desousa@unb.br

**Abstract**—Drivers' behavior in traffic is a determining factor for the rate of accidents on roads and highways. This paper presents the design of an intelligent IoT system capable of inferring and warning about road traffic risks and danger zones, based on data obtained from the vehicles and their drivers mobile phones, thus helping to avoid accidents and seeking to preserve the lives of the passengers. The proposed approach is to collect vehicle telemetry data and mobile phone sensors data through an IoT network and then to analyze the driver's behavior while driving, along with data from the environment. The results of the inference serve to alert drivers about incidents in their trajectory as well as to provide feedback on how they are driving. The proposal is validated using a developed prototype to test its data collection and inference features in a small scale experiment.

**Index Terms**—Internet of Things; Vehicular networks; Driving behavior; Inference; OBD-II; Android.

## I. INTRODUCTION

THE WAY drivers behave in traffic is a determining factor for high accident rates on roads and highways. In 2018, there were 69114 serious accidents on Brazilian highways [1]. In such context [2], by analyzing the driver's profile, it is possible to analyze the phenomenon and create mechanisms to positively influence driver behavior, thus making the routes safer and energy use more efficient.

Over time, technologies, such as smartphones, have become easily accessible to the population and have aided drivers. According to FGV-SP [3], the number of smartphones in Brazil already exceeds 230 million. Since most of these mobile phone models have sensors, such as GPS (Global Positioning System), accelerometer, gyroscope, 3-axis magnetometer (lateral, longitudinal and vertical or in coordinates  $x$ ,  $y$  and  $z$ ), we can use their data, as indicated by other studies, such as [4] or [5], to contribute to the driver behavior study.

Internet of Things (IoT) is a paradigm that combines aspects and technologies of different approaches: ubiquitous computing, communication protocols and technologies, sensors and actuators, composing a system in which the real world and the digital world interact symbiotically [6]. The IoT connected devices installed base comprised around 23.14 billion devices

in 2018 and it is projected to increase to 75.44 billion ones worldwide by 2025 [7].

The increasing number of internet-connected objects ranging from cell phones to air conditioners is a compelling force for a more comprehensive study of how such devices connect. The ease to share information among IoT devices open up a new environment for different uses, where objects with traditional use can turn into objects with a certain intelligence, as shown in [8].

The IoT concept is based on data sharing between several different devices, be they vehicles or, in their simplest form, smartphones. Thus, shared data can be treated in a way that generates different interpretations and providing significant indicators to, for example, influence users behavior [9]. Then, the device can be part of several IoT networks and with intelligence to infer actions.

For the proposed project development, firstly it was necessary to fully understand the IoT concept, which is used for new technology development and it is the assumption on which the Smart Drive project is based, developed and discussed in this article.

Like all new technologies, the IoT development also faces several challenges. One of main topics addressed in our IoT study is network security. The need for in-depth study of this topic is observed when analyzing the steps necessary for system operation, i.e. information sharing may be subject to malicious actions, which will compromise the network operation and even the user privacy. Some methods are employed to mitigate this problem, as explained in [10]. This topic was approached with its necessary applications in the project being registered in this paper.

The rest of this article is divided into four parts. Section II briefly highlights the main related works. Section III presents the Smart Drive project proposal as well as an overview of what was developed. Section IV encompasses the project development and implementation. Section V focuses on the main contribution of this paper regarding inferences of driver behavior. Finally, Section VI presents the conclusions about the validation of the proposal and comments on future work.

## II. RELATED WORKS

In order to start the proposed project, it was necessary to first investigate some related works, i.e. already published articles that had some ideas correlated with the objectives here exposed. Therefore, the following articles were analyzed:

- Driver Behavior Profiling Using Smartphones: A Low-Cost Platform for Driver Monitoring [11]: In this article, it is analyzed how smartphone sensors can be used to identify maneuvers. SenseFleet is proposed, a steering profile platform that is able to detect risky events direction; and
- Driver Profile Analysis: Event Detection through Smartphones and Machine Learning [2]: This article conducts an investigation with different sensors, present in an Android smartphone, and different classification algorithms, in order to evaluate which sensor set/method allows classification with greater accuracy. The results show that specific combinations of sensors and intelligent methods allow to improve rate performance.

It is important to note that other articles were also considered for the realization of the project and are cited hereafter.

## III. METHODOLOGY

For the project development, it was necessary to define what should be inferred and what should be returned to the user.

The project scope was based on the environment in which the car will be present, as well as its movement, given data sampling of several routes performed. So, it was possible to predict user behavior and to correlate data with streets that the car would go through. Thus, user's mobile phone can inform the best route as well as possible dangers in its trajectory.

The referred paper [12] is very useful regarding the functionality developed in the project, as well as the data to be inferred and corresponding interpretations. Also interesting is [2], which provides a good basis for the driver profile analysis.

### A. Objectives

Six specific objectives, that will provide a project progress vision, have been defined:

- 1) Collect telemetry data from vehicular computer. Such data will be captured from OBD-II (On-board diagnostics) device, as indicated in [5];
- 2) Collect data from user Android device sensors, as indicated in [13] and [14];
- 3) Develop mobile application gateway to receive collected data and transmit to server called Smart Driver;
- 4) Develop secure server, using Hypertext Transfer Protocol Secure protocol, with auto registration function to receive collected data and transmit it to application;
- 5) Execute inferences, both at application and server layers, using collected data, to identify driver behavior, as well as to send geographic information on the risk areas;
- 6) Develop application for users administration, access to collected data and inference results.

### B. Project phases

The architecture development has occurred in five phases, detailed as follows: in phase 1 data were collected from an OBD-II device, connected to the car, and the individual's smartphone; phase 2 consists of access to Smart Driver platform, from an application installed on smartphone; in phase 3 each user direction characteristics are verified by inference; Phase 4 consists of indications to the user about incidents generated from his driving and information about safety of certain routes; and phase 5 provides user administration interface, data and inferred information. Figure 1 shows, in an illustrative way, the phases related above.

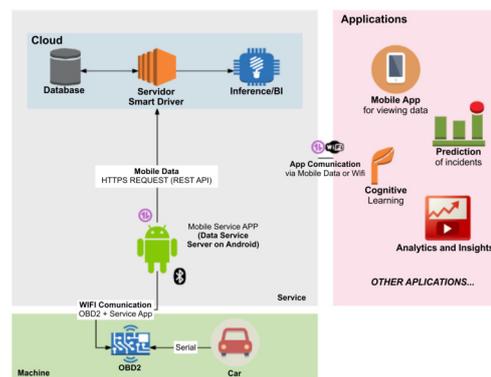


Figure 1. The project general architecture

### C. Sensors for an Android device

Some data have been set to be used from the user's Android device. These data are considered from the cell phone capability of acquiring them, for that, the work presented in [13] was useful. The data are quoted below:

- accelerometer - abrupt acceleration or deceleration can be inferred according to the acceleration vector provided by the sensor;
- GPS - provides the speed (m/s), making it possible to compare this value with the allowed track speed;
- orientation - according to the magnetometer and the gravity sensor, azimuth ( $-\pi$ ,  $\pi$ ) is obtained in radians. The change rate at the steering wheel is found by calculating the change after two subsequent samples, giving the idea of how sharp is the car turning.

The Android operating system was chosen as basis to this research because it is installed on almost seven times more devices than the iOS operating system in Brazil.

## IV. PROPOSED SOLUTION

This topic reports the development of proposed application for recording incidents and events while driving. The application module is responsible for receiving information generated by user and present it to the UIOT, a middleware responsible for data storage and sharing, developed and maintained by IoT research team, at the University of Brasilia. The tools used to develop the application module were Java, Android Studio IDE and Google Firebase.

### A. SmartDriver Platform

The SmartDriver platform development objective was to create a secure service for communication of stealthy data about user location, requiring that this service being scalable and highly available through the "Raise Middleware" use. Another important goal was the development of the user administration interface, the data and the plotting of the routes and cluster of incidents as heat zones on the map.

In Figure 2 is possible to visualize the reference region heat map, where heat points (red scales), present regions in which there were incidents, such as, abrupt acceleration and stops.

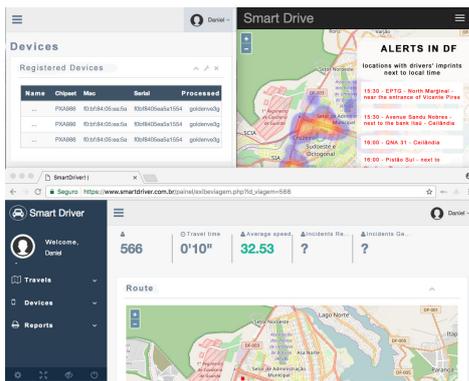


Figure 2. SmartDriver Platform

In this platform, in addition to presenting the risky areas on the roads, with their respective history of alerts, it is also possible to register new users. In order to do that, the Log in / Register window must be accessed to allow user will to set the requested email and password.

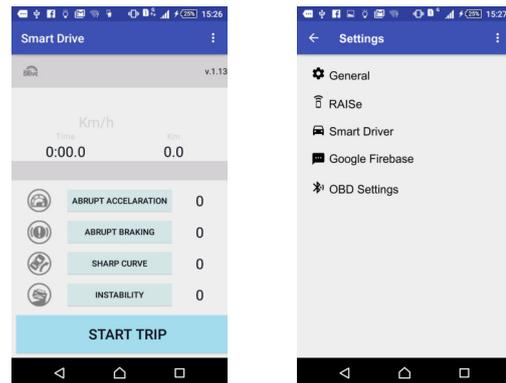
### B. Proposed mobile application

The application for capturing events has two main screens: Home and Settings.

The application starting screen is shown in Figure 3a. The user's trip will be initiated by clicking the Start Trip option. In this screen, the user's car speed will be displayed as well as the traveled time, the distance traveled and all events like abrupt acceleration, abrupt breaking, sharp curve and instability.

When starting the trip, in addition to tracking possible events while driving, user can report third-party events by clicking on event name buttons, as shown in Figure 3a.

Finally, the user will also have access to application settings, as shown in Figure 3b. In this menu, it is possible to access general application settings; connectivity function with Raise; connectivity settings with Smart Driver server; OBD-II device settings and the Firebase function that stores the updated Firebase token value, which identifies each device/user. This value will be used by the inference to send the results (data to be presented to the user) directly to device, which will be read by application and presented to user.



(a) Main screen

(b) Settings screen

Figure 3. Application screens

### C. Self-registration

The self-registration mechanism was developed as a way to facilitate the entry and management of devices in IOT networks, allowing sensors and actuators, aware of their context, to enter autonomously and securely in an environment capable of receiving massive volumes of data and control actuators safely. As explained in [15], the middleware responsible for receiving and processing data in cloud has two main components, the REST API Approach for IoT Services (RAISe) and the User Interface Management System (UIMS). RAISe is the web services interface responsible for responding to client requests, storing data provided by these clients. UIMS is the visual interface through which a user can consult data manipulated by the middleware.

The solution implements a complete self-registering architecture for the Smart Device from sending basic device data, such as serial number, MAC address and other identification data, thus forming a unique composite primary key for any device that enters into that network. After registration, the device can be associated with one or more users, keeping track of who handled the device during its life cycle. Thus, the sensor data recording is associated to user, promoting individual definition of their steering profile.

The entire information transit process between the device and the self-registering middleware is performed through a Secure Sockets Layer (SSL) connection. In addition, sender authenticity is verified based on a token generated randomly and periodically by the middleware and sent to the device after its registration process. Thus, in order to allow the call of service from a device, it must send an authentication token which must be within its validity.

Due to need for additional processing, the communication architecture between device and middleware is asynchronous. Thus, although the device makes repeated calls to the service to perform the data sending, the inference answer is performed only after processing data step, with message sending through an operating system API call. This solution was adopted because the constant data sending process from client requires a real-time response from the server.

## V. INFERENCE

The inference process is the base for the *insights* obtained from the collected raw data of the IoT devices. In the proposed model, part of the data processing and analysis is performed in the device itself, such as the identification of sharp braking and curves, and part is performed in a centralized way, mainly heuristics that depend on collective knowledge, that is, involving joint analysis of data coming from different devices.

Heuristics involving collective knowledge are precisely the main contribution of this project. The centralized analysis allows not only heuristics creation involving large volume of data, due to the difference of processing and storage capacities in relation to IoT devices, as it creates a bidirectional channel of communication, allowing the collectively generated *insights* to arrive at each of the individual devices. It is important to emphasize that, although several studies propose Vehicle-to-Vehicle communication models (V2V communication), the cloud-based communication strategy is more adequate at the present moment, since, while the mobile data communication technologies are more consolidated and provide good coverage and high speeds, V2V technologies are still in the early stages of deployment.

### A. Clustering and Alerts

From the collected data, the incident point clustering is performed to calculate the region hazard index and, thus, define the hazardous areas. The clustering groups points that they are at a distance of three meters from each other; this distance is necessary because it is very unlikely that two incidents will be marked exactly in the same geographic coordinate. Thus, the clustering process assists in the identification of areas where there is incident point concentration.

It was considered as an alternative the fact that, from the existing data sample, an inference was made to prove the proposed IoT architecture intelligence layer from the service layer, as well as to provide practical and efficient results regarding solution benefits. In this way, an automatic inference was developed, based on data from the Android device sensors, which analyzes if the speed sent to the Smartdrive server (middleware) is greater than track speed by 10%, if so, using Firebase, the server sends an "Over Speed" alert to the application and to the Android device's messaging system by means of a standard message class that logs all sent alerts as shown in Figure 4a.

Speed alerts are generated by regulatory speed capturing of route on which user is traveling. This is done through a geographic consult to the Openstreetmap geographic database, which was imported into the inference layer local database, a PostgreSQL DBMS with the postGIS extension.

User sends his geographical coordinate every 0.5 seconds; others sensor data are activated if the system identifies that the coordinate sent by user is within 50 meters of a hazardous area. Thus, a notification is sent to user informing that he is approaching a hazardous area as shown in Figure 4b.

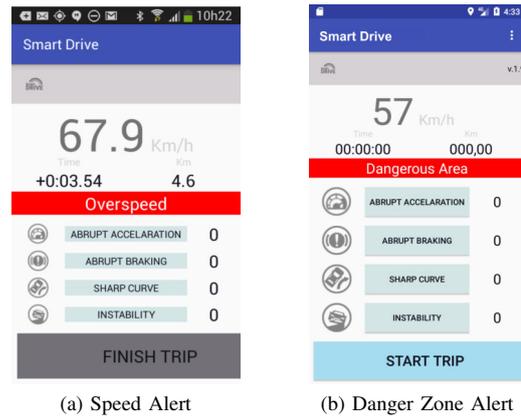


Figure 4. Alert screens

### B. Execution Layer Inference

Inferences executed only at service layer would not characterize an IoT solution. Given the application on smartphone could not be considered a smart object whether it did not have embedded intelligence, it was decided to make inferences about driver behavior directly in execution layer, so that its results can be sent to server in followed or simultaneously, depending on packet sending time configuration.

Again, to prove initial hypothesis in the work of the available data sample, we opted to identify two physical phenomena: abrupt acceleration and abrupt braking, from reading the three axes (x, y, and z) of the Android device's linear acceleration sensor.

Using simple logic without noise (Kalman), and as reference, the gravity acceleration (9.8 m/s) and threshold defined in [16] [17], the following equation has been set up to alert whether values exceed thresholds set for abrupt braking and acceleration (1): a is accelerometer X axis sensor measurement, b is Y axis measurement and c the measuring of Z axis.

$$0.4g < \sqrt{a^2 + b^2 + c^2} \quad (1)$$

Another inference implemented in execution layer is to use vehicular computer data through OBD-II protocol, in which, whenever vehicular sensor acceleration is greater than 1, both application and trip will start automatically. Besides that, the trip will be automatically interrupted when speed and engine rotation are both equal to ZERO.

## VI. CONCLUSION

With goal of validating the system and scaling a cloud solution capable of meeting the demand of a large number of users, we invited 50 volunteers to use our software for a period of 30 days, with an average usage of three hours per day and we obtained the following results:

- 1) 24 users (50% of the sample) did not have EML327 interface in their car;
- 2) 5 users (10% of the sample) had communication problems with the OBD-II adapter;

- 3) 3 users (6% of the sample) have managed to install the OBD-II adapter but did not use the application at suggested frequency; and
- 4) 18 users (36% of the sample) used the application with suggested daily frequency.

Analyzing the last group of 18 users, we have the following statistics:

- 1) smartphone sensors collected a mean of 160 Bytes/s;
- 2) OBD-II interface captured a mean volume of 80 Bytes/s.

Therefore, the average data rate generated by the solution was 250 Bytes per second. With these values, we can conclude that in a journey of one hour, each mobile phone will need to transmit to the middleware 900 MB of data collected.

After the data collection sent by the group of volunteers, we arrived at the following conclusions:

- 1) Some manufacturers have not yet fully adhered to the OBD-II protocol, either for reasons of industrial secrecy or for adhering to underlying standards as proposed by the European Union and some Asian countries.
- 2) The generated data volume by each mobile phone is very high. This fact is leading us to research solutions where the mobile phone does not send all the data to central middleware. Taking advantage of today's phones large processing capacity, we state as future work a pre-processing at the edge, sending to the middleware only alert information and summary statistical data.
- 3) The client side processing and data analysis to identify individual events, such as sharp braking and presence of curves, seems to be a good alternative to reduce the need for processing capacity in central server.
- 4) The proposal heuristics involving collective knowledge creates a bidirectional channel of communication, allowing the collectively generated insights to arrive at each of the individual devices.
- 5) At this moment, the cloud-based communication strategy seems to be more adequate to provide communication among vehicles, mainly because V2V technologies are still in the early stages of deployment.

The security aspects or GDPR were not addressed in depth in this work since the goal was to evaluate the proposed platform operational capability. A deeper analysis of these security and privacy issues and the conduction of new tests with a higher number of users are considered for future work.

#### ACKNOWLEDGMENT

This research work is supported by the Brazilian research and innovation Agencies CAPES (Grants 23038.007604/2014-69 FORTE and 88887.115692/2016-00), CNPq (Grant 465741/2014-2 INCT-CyberSecurity), and FAPDF (Grants 0193.001366/2016 UIoT and 0193.001365/2016 SSDDC), as well as the LATITUDE/UnB Laboratory (Grant 23106.099441/2016-43 SDN), the Ministry of the Economy (Grants 005/2016 DIPLA, 011/2016 SEST and 083/2016 ENAP), the Institutional Security Office of the Presidency of the Republic of Brazil (Grant 002/2017), and the IEEE VTS Centro-Norte Brasil Chapter.

#### REFERENCES

- [1] Brasil. (2019) Portal oficial de notícias da Polícia Rodoviária Federal: Balanço PRF 2018. [Online]. Available: <https://www.prf.gov.br/agencia/prf-registra-diminuicao-no-numero-de-acidentes-e-mortes-nas-rodovias-federais-em-2018>
- [2] J. Ferreira Júnior and G. Pessin, "Análise de perfil de motoristas: Detecção de eventos por meio de smartphones e aprendizado de máquina," in *Anais do WOCES 2016 Workshop de Comunicação em Sistemas Embarcados Críticos*, 2016, pp. 76–85.
- [3] FGV-SP. (2019) Pesquisa anual do uso de TI da Fundação Getúlio Vargas-SP. [Online]. Available: <https://easp.fgv.br/ensinoconhecimento/centros/cia/pesquisa>
- [4] S. R. Muramudalige and H. D. Bandara, "Demo: Cloud-based vehicular data analytics platform," in *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services Companion*, ser. MobiSys '16 Companion. New York, NY, USA: ACM, 2016. doi: 10.1145/2938559.2948849. ISBN 978-1-4503-4416-6 pp. 1–1. [Online]. Available: <http://doi.acm.org/10.1145/2938559.2948849>
- [5] M. Amarasinghe, S. Kottogoda, A. L. Arachchi, S. Muramudalige, H. M. N. Dilum Bandara, and A. Azeez, "Cloud-based driver monitoring and vehicle diagnostic with obd2 telematics," in *2015 IEEE International Conference on Electro/Information Technology (EIT)*, May 2015. doi: 10.1109/EIT.2015.7293433. ISSN 2154-0373 pp. 505–510.
- [6] E. Borgia, "The Internet of Things vision: Key features, applications and open issues," *Computer Communications*, vol. 54, pp. 1–31, Dec. 2014. [Online]. Available: <http://dx.doi.org/10.1016/j.comcom.2014.09.008>
- [7] S. R. Department. (2016) Internet of things (iot) connected devices installed base worldwide from 2015 to 2025 (in billions). [Online]. Available: <https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/>
- [8] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of things: A survey on enabling technologies, protocols, and applications," *IEEE Communications Surveys Tutorials*, vol. 17, no. 4, pp. 2347–2376, Fourthquarter 2015. doi: 10.1109/COMST.2015.2444095
- [9] B. Xiao, R. Rahmani, Yuhong Li, D. Gillblad, and T. Kanter, "Intelligent data-intensive iot: A survey," in *2016 2nd IEEE International Conference on Computer and Communications (ICCC)*, Oct 2016. doi: 10.1109/CompComm.2016.7925122 pp. 2362–2368.
- [10] T. Xu, J. B. Wendt, and M. Potkonjak, "Security of iot systems: Design challenges and opportunities," in *2014 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, Nov 2014. doi: 10.1109/ICCAD.2014.7001385. ISSN 1092-3152 pp. 417–423.
- [11] G. Castignani, T. Dermann, R. Frank, and T. Engel, "Driver behavior profiling using smartphones: A low-cost platform for driver monitoring," *IEEE Intelligent Transportation Systems Magazine*, vol. 7, no. 1, pp. 91–102, Spring 2015.
- [12] G. Castignani, T. Dermann, R. Frank, and T. Engel, "Driver behavior profiling using smartphones: A low-cost platform for driver monitoring," *IEEE Intelligent Transportation Systems Magazine*, vol. 7, no. 1, pp. 91–102, 2015.
- [13] V. Astarita, G. Guido, D. Mongelli, and V. P. Giofrè, "A co-operative methodology to estimate car fuel consumption by using smartphone sensors," *Transport*, vol. 30, no. 3, pp. 307–311, 2015.
- [14] B. P. Puig, "Smartphones for smart driving: a proof of concept," *unpublished master's thesis for master's degree, Universitat Politècnica de Catalunya, Barcelona*, 2013.
- [15] C. C. d. M. Silva, F. L. d. Caldas, F. D. Machado, F. L. Mendonça, and R. T. de Sousa Júnior, "Proposta de auto-registro de serviços pelos dispositivos em ambientes de iot," *34º Simpósio Brasileiro de Telecomunicações e Processamento de Sinais*, 2016.
- [16] Hongyang Zhao, Huan Zhou, Canfeng Chen, and J. Chen, "Join driving: A smart phone-based driving behavior evaluation system," in *2013 IEEE Global Communications Conference (GLOBECOM)*, Dec 2013. doi: 10.1109/GLOCOM.2013.6831046. ISSN 1930-529X pp. 48–53.
- [17] J. Paefgen, F. Kehr, Y. Zhai, and F. Michahelles, "Driving behavior analysis with smartphones: Insights from a controlled field study," in *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia*, ser. MUM '12. New York, NY, USA: ACM, 2012. doi: 10.1145/2406367.2406412. ISBN 978-1-4503-1815-0 pp. 36:1–36:8. [Online]. Available: <http://doi.acm.org/10.1145/2406367.2406412>



## Smart Urban Design Space

Philipp Skowron  
Leipzig University, Germany  
Email: skowron@wifa.uni-leipzig.de

Michael Aleithe  
Leipzig University, Germany  
Email: aleithe@wifa.uni-leipzig.de

Susanne Wallrafen  
Sozial-Holding der Stadt  
Mönchengladbach GmbH,  
Germany  
Email: s.wallrafen@sozial-holding.de

Marvin Hubl  
University of Hohenheim,  
Germany  
Email: marvin.hubl@uni-hohenheim.de

Julian Fietkau  
Universität der Bundeswehr  
München, Germany  
Email: julian.fietkau@unibw.de

Bogdan Franczyk  
Wroclaw University of Economics,  
Poland  
Email: bogdan.franczyk@ue.wroc.pl

**Abstract**—The irreversible process of demographic change, especially in Germany, leads to numerous challenges. According to this, research has to face the task to integrate the constantly ageing population into the urban and public space in such a way that there are as few barriers as possible. With the support of digitalization, so-called smart urban objects are being designed in order to do make integration, so that people and the available technology can be used most efficiently. A special ontology has been developed to meet this demand.

### I. INTRODUCTION

The demographic change of a permanently ageing population has become a globally visible phenomenon. Particularly in Germany, the population will be considerably older in the future than it is at present. According to [1], every third person will be older than 65 years of age by 2060. Corresponding with the tendency of a permanently ageing population goes the fact of changing needs and in daily life. In the era of the inevitable digitalization and in particular the *Internet of Things* (IoT), the challenge is to what extent digitalization can improve daily life for these ageing population. Accordingly, the concept of providing the urban space with so-called *smart urban objects* (SUOs) [2] is being pursued to increase the participation of elderly people by digitalization. These SUOs are elements of the urban environment, e.g. lights, information boards and benches, which are connected to a digital information space and allow for implicit or explicit interaction. The desired goal is to increase the feeling of security on urban environment by personalization of these objects. Some of these SUOs are described in detail in [2], [3], [4], [5] and [6]. The focus of this research is the intersection between the behavior of elderly people, currently referred to as *Ambient Assisted Living* (AAL), and *Smart City*. The final focus of this paper is to provide an ontology for classifying these SUOs so that both the technical aspects as well as the aspects of the AAL are considered.

### II. MOTIVATION AND RESEARCH QUESTION

The increase in barrier-free accessibility, especially for older persons, will be achieved with the support of SUOs. In order to enable a categorization of these objects, an ontology is required which takes both technical aspects and the view of public health and AAL into account. Based on this kind of ontology, designers of SUOs can consider all aspects mentioned to achieve maximum efficiency of these objects in the later context. In order to sufficiently answer this motivation, following research question is posed, which is the central issue of this article.

*How does a taxonomy for the design of SUOs have to be constructed in order to sufficiently consider aspects of Public Health and AAL as well as the technical perspective, so that a maximum increase of barrier-free accessibility is already addressed during the design process?*

### III. RELATED WORK

At this point, approaches and solutions are described and analyzed in terms of the way they answer the research question of this article. Basically, ontologies exist on the one hand in the field of Smart City and on the other hand in the field of so-called *Public Health*. At this point, both directions will be analyzed in depth and compared with each other, though the research question here characterizes exactly the intersection between these two directions.

In [7] an ontology in the area of Public Health is described, which characterizes in particular the direct situation in the hospital. Here so-called *medical classes* and *medical activations* exists. The former include specific diseases, symptoms, therapies, roles and departments in the hospital. The activations subsequently serve to bring these medical classes together in a meaningful relationship and thus describe the applications in the field of Public Health. An ontology-based approach in public health with the support of a *geographic information system* (GIS) is discussed in [8].

The ontology is used for the fusion of data from social and health related issues. Nevertheless, the GIS is the primary focus of the description, and ontology is only used as a tool. So therefore is no further discussion of it. In the contribution of [9] a set of different ontologies is presented, which should support designers in the development of so called AAL and those services. In detail, *actors*, *spaces* and *devices* are modeled and linked so that concrete AAL-elements can be described that have been used within the present study. Overall, this type of modeling is very complex and still has no generic character, meaning that any further use is crucial. A framework for managing the current state as well as the users profile information extracted from the internet and the mobile context is illustrated in [10]. This so called *Next Generation Network* (NGN) is an ontology for modeling typical users of AAL-services. But these services are only user centric and have no relation to technical issues. Also the platform in [11] offers assistance in communication and information acquisition by providing personalized and context-aware AAL-services. Therefore an ontology is used where users are the central aspect of the platform. Furthermore this ontology enables a historical view of the users changing characteristics and environment. In view of this explanation, only the user behavior is addressed without encompassing the technical factors. Also in [12] an ontology for structuring daily living activities of users is depicted, whereby a stronger focus is placed on the underlying aspect of AAL and thus on elderly persons. In contrast, the ontology in [13] discusses the technical aspects in terms of best practice for building automation devices and functions and how these underlying models are structured especially in the area of AAL. But in this case there is only a technical view without inclusion of users perspective.

In contrast to solutions of Public Health and AAL, there are some approaches from the Smart City context. These are presented in the following. This Smart City context is characterized by data collected from various distributed systems. Purposing these task in [14] the so called *Semantic Web* is used for designing a new Smart City ontology. The primary focus is to address the interoperability among the different systems and frameworks for describing Smart City objects. In [15] is an analysis about the impact of Smart City applications observed in the field of energy and transport. Besides [15] describes [16] an ontology to describe the entire Smart City domain. In [17] this description is extended for IoT-based applications. Nevertheless, [14], [15], [16] and [17] all have a strong technical focus and do not mind the user-centered perspective.

In addition to the number of ontologies mentioned so far, a so-called *Design Space* is described in [18] which enables the characterization and categorization of UI-based elements in the development of applications. This idea would require continuous expansion to include the sensors and applications of the IoT arising.

In summary, a wide spectrum of previous ontologies were presented. These addresses on the one hand the areas of

Public Health and AAL and on the other hand the topic of Smart City. The former ontologies have a strong user-centric focus and the latter are technically very pronounced. However, there is no solution among all approaches that represents a sufficient mix to satisfy the related research question of this article. In addition, the investigated solutions indicate that the aspect of interconnecting the underlying data structure is becoming increasingly important. As a result, this aspect would also have to be integrated more into the ontologies used in this context.

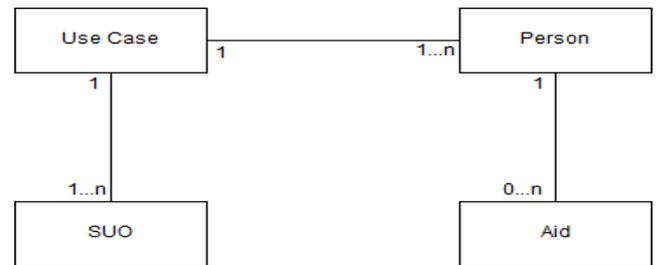


Fig 1. This picture illustrates the entity relationship diagram of the basic relation between smart urban objects (SUOs), the appropriate use case, involved persons and the personalized purpose (aid).

#### IV. DESCRIPTION OF SMART URBAN DESIGN SPACE

This chapter introduces the so-called *Smart Urban Design Space* (SUDS). Such a taxonomy meets the above-mentioned full range of criteria in terms of technical aspects, AAL and Public Health. A fundamental idea of this SUDS is the networking of the separate criteria. The basic context is represented graphically in Fig. 1, whereby each use case can be supported by at least one or more SUOs, which are used by at least one or more persons. In order for the SUOs to be used by the persons per use case, it may be necessary to provide additional assistance, which is continuously referred to as aid.

Against this background, a use case is the concrete scenario in which the elderly person(s) can use the digital support outdoors (outside buildings). Concrete examples in this context are an adaptive lighting system of the area to be walked in during a walkway, an adaptive park bench, which adapts to the individual sitting height of the respective person as well as intelligent information spotlights, which provide personalized information of the urban space to be visited. These examples are presented in detail in chapter 5.

Within the SUDS, the three criteria SUO, Aid, and Person exist for each use case, with their corresponding subordinate properties. In this regard, an overview of the entire taxonomy is shown in Fig. 2. A person has so-called competencies, which are continuously referred to as *skills*. These include *speaking*, *seeing*, *hearing*, *cognitive* skills such as easy logical thinking and *movement*, which in this case refers to walking without aids. The SUO contains the five criteria *actuator*, *sensor*, *parallelization*, *personalization* and *interaction sensor*. The *interaction sensor* describes which human

sense for an interaction of the SUO is required. It distinguishes between *seeing*, *hearing* and *haptic* handling such as using a touch pad. In addition to operating sensors, there is also the criterion of technical sensors, which is referred to merely as *sensors* within this taxonomy. There are *mechanical*, *piezoelectric*, *capacitive*, *inductive*, *optical*, *magnetic* and *signal-based* practices. The latter symbolize the provision of information by an external information source. Similar to technical sensor technology, the *actuator* also distinguishes between *mechanical*, *signal-based*, *optical*, *thermal* and *acoustic* variants. *Personalization* classifies the SUO according to whether each *individual* person is addressed individually, whether a group of people is addressed (*cluster*) or whether no individual personalization (*general*) is satisfied. In this context, there is also the criterion of *parallelization*, whether the SUO differentiates only *single-user* or *multi-user* in the respective use case. Similar to the SUO, the aid has a shortened set of criteria. The *interaction sensor*, *actuator* and *sensor* are used, with the latter describing the technical perspective. The characteristics of these criteria are analogous to those of the SUO.

V. CASE STUDY “URBANLIFE+”

In the research project *UrbanLife+*, the autonomy and participation of senior citizens in urban areas is explored in such a way that they can be increased. For this purpose, urban objects in Mönchengladbach are to be transformed into SUOs with the help of innovative human-technology interaction approaches, which provide senior citizens with technical support in line with their needs and enable them to move around the city safely [2] [5]. Three use cases are pre-

sented for these addressed solutions, which are then classified in the SUDS. These use cases are Adaptive Lighting, Adaptive Park Bench and the Information Radiators. In the following these are explained briefly and the classification in the SUDS is discussed individually. Overall it is represented in Fig. 2.

A. Adaptive Lighting System

The system of the Adaptive Lighting improves the feeling of safety on elderly people especially in dark areas at night by personalized and position-dependent variation of intensity and/or color of the light [1] [4].

B. Adaptive Park Bench

Adaptive Park benches are a kind of smart seats, that can adjust to individual anthropometric measures of people. Thereby the usability of the seats is enhance which in turn also enhances safe usage. Particularly older people face severe problems in sitting down and standing up at common seats. The reason is that, the gap between standover and the height of the seat surface imposes trouble when the older people have weakened leg muscles, impaired balance or general difficulties in bending their knees. For this reason the seat surface of the adaptive park bench can lift up to the standover of a pedestrian, which actively supports in sitting down and standing up. For ergonomic sitting the seat surface will be adjusted to the sitting person’s popliteal height. More technical details are described in [2] [19].

C. Information Radiators

Information Radiators are a class of devices capable of displaying dynamic information while installed at a static

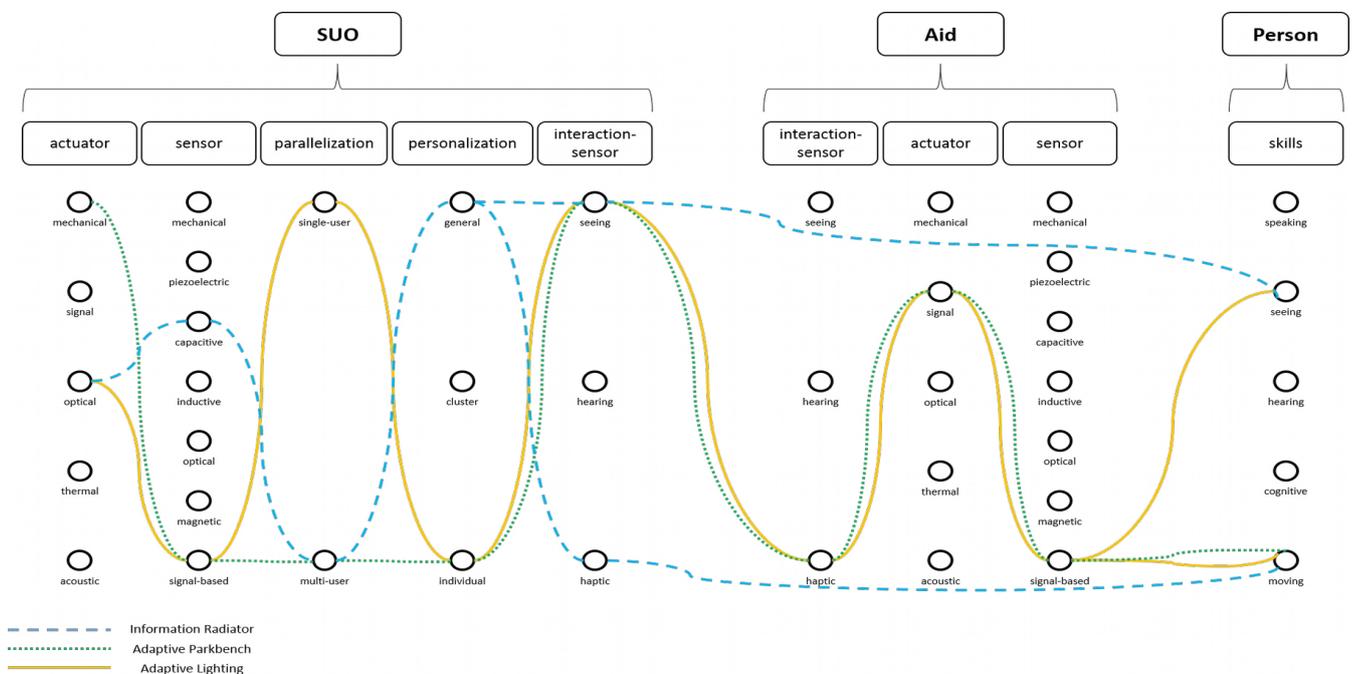


Fig 2. This figure is the representation of the so-called *Smart Urban Design Space* (SUDS). It shows the basic categories SUO, Aid and Person, with the available properties. By combining the properties of these categories addressed by the corresponding use case, a visualization similar to a dendrogram is generated. Furthermore in this figure, the three use cases are arranged in the SUDS. These use cases include the *Adaptive Lighting*, the *Adaptive Park Bench* and the *Information Radiator*, each of which is marked in the legend.

position in a public or semi-public environment. They can range from large interactive screens which can be used via touch, to small low-power devices equipped with low-resolution LED displays. The informational content they show is related to the local context. This includes, but is not limited to, offerings by commercial and noncommercial actors in the vicinity (such as stores, restaurants, cinemas, community centers, sports clubs, etc). The concept is discussed in more detail in [20].

## VI. DISCUSSION

As a result of the taxonomy of the SUDS developed in this paper, a rather interdisciplinary classification of the so-called SUOs has been successfully achieved, without getting stuck in technical details in this context, nor without too one-sided a view of social criteria affecting the user. In this context, the previous SUOs were specifically classified in the SUDS (see Figure 3). The SUOs classified so far include Adaptive Lighting, the Adaptive Park Bench and the Information Radiators. In addition to these existing SUOs, there is also the possibility of continuously classifying new ones in order to visualize the essential aspects of the local field of knowledge. The classification in this taxonomy (Figure 3) shows that more or less all SUOs have similar characteristics regarding their categories. For example, in relation to the interaction sensor, which is only haptically or optically pronounced in all previous objects. Consequently, an essential motivation for further SUOs would be to include acoustic signals in order to increase the intersection between the technical and personal skills.

## VII. CONCLUSION AND OUTLOOK

Concerning the research question within this article, a taxonomy called SUDS was constructed which merges the required aspects of AAL, Public Health and technical aspects and makes them usable for integrating so-called SUOs.

In the future, potentially beneficial SUOs could be determined and designed with the support of the SUDS, which do justice to the aspects of AAL and Public Health without violating the technical conditions.

## ACKNOWLEDGEMENT

This work was fully conducted in the scope of the research project *UrbanLife+* (16SV7442), funded by the German Ministry of Education and Research.

## REFERENCES

- [1] <https://de.statista.com/statistik/daten/studie/71539/umfrage/bevoelkerung-in-deutschland-nach-altersgruppen/>, last accessed: 2019-03-11
- [2] Hubl, M., Skowron, P., Aleithe, M.: Towards a Supportive City with Smart Urban Objects in the Internet of Things: The Case of Adaptive Park Bench and Adaptive Lights. In: Position Papers of the 2018 Federated Conference on Computer Science and Information Systems (FedCSIS). Annals of Computer Science and Information Systems (ACSIS) 16. Maria Ganzha, Leszek A. Maciaszek, Marcin Paprzycki (eds.), pp. 51-58, (2018). doi:10.15439/2018F118
- [3] Kötteritzsch, A., Koch, M., Wallrafen, S.: Expand Your Comfort Zone! Smart Urban Objects to Promote Safety in Public Spaces for Older Adults. In: Adjunct Proceedings of UbiComp 2016, ACM Press, (2018) doi:10.1145/2968219.2968418
- [4] Aleithe, M., Skowron, P., Franczyk, B., Sommer, B.: Data modeling of smart urban object networks. In: Proceedings of the International Conference on Web Intelligence (WI '17), pp. 1104-1109 (2017). doi:10.1145/3106426.3117759
- [5] Aleithe, M., Skowron, P., Schöne, E., Franczyk, B.: Adaptive Lighting System as a Smart Urban Object. In: Communication Papers of the 2018 Federated Conference on Computer Science and Information Systems (FedCSIS 2018). Annals of Computer Science and Information Systems (ACSIS) 17, Maria Ganzha, Leszek A. Maciaszek, Marcin Paprzycki (eds.), pp. 145-149, (2018). doi:10.15439/2018F30
- [6] <https://www.urbanlifeplus.de/>, last accessed: 2019-02-26
- [7] Abinaya, Kumar, V., Swathika: Ontology Based Public Healthcare System in Internet of Things (IoT). In: Procedia Computer Science 50, pp. 99-102, (2015). doi:10.1016/j.procs.2015.04.067
- [8] Gür, N., Sanchez, L. D., Kauppinen, T.: GI Systems for Public Health with an Ontology Based Approach. In: Proceedings of the AGILE/2012 International Conference on Geographic Information Science, Avignon, pp. 86-91 (2012). ISBN: 978-90-816960-0-5
- [9] Mocholi, J. B., Sala, P., Fernandez-Llatas, C., Naranjo, J. C.: Ontology for Modeling Interaction in Ambient Assisted Living Environments. In: XII Mediterranean Conference on Medical and Biological Engineering and Computing (MEDICON 2010), pp. 655-658 (2010). doi:10.1007/978-3-642-13039-7\_165
- [10] Moreno, P. A., Hernando, M. E., Gomez, E. J.: AALUMO: A User Model Ontology for Ambient Assisted Living Services Supported in Next-Generation Networks. In: XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013, pp. 1217-1220 (2013). doi:https://doi.org/10.1007/978-3-319-00846-2\_301
- [11] Fredrich, C., Kuijs, H., Reich, C.: An Ontology for User Profile Modeling in the Field of Ambient Assisted Living. In: SERVICE COMPUTATION 2014 : The Sixth International Conferences on Advanced Service Computing. (2014). ISBN: 978-1-61208-337-7
- [12] Woznowski, P. R., Tonkin, E. L., Flach, P. A.: Activities of Daily Living Ontology for Ubiquitous Systems: Development and Evaluation. In: Sensors 2018, 18, 2361. (2018) doi:10.3390/s18072361
- [13] Butzin, B., Golatowski, F., Timmermann, D.: A survey on information modeling and ontologies in building automation. In: Conference: IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society. (2017). doi:10.1109/IECON.2017.8217514
- [14] Abid, T., zarzour, H., Laouar, M. r., Khadir, M. T.: Towards a smart city ontology. In: Conference: 2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA). (2016). doi:10.1109/AICCSA.2016.7945823
- [15] Komninos, N., Bratsas, C., Kakderi, C., Tsarchopoulos, P.: Smart city ontologies: Improving the effectiveness of smart city applications. In: Journal of Smart Cities 1(1), pp. 1-16 . (2015). doi:10.18063/JSC.2015.01.001
- [16] Ramaprasad, A., Sanchez-Ortiz, A., Syn, T.: A Unified Definition of a Smart City. In: EGOV 2017, LNCS 10428, pp. 13-24, (2017). doi:10.1007/978-3-319-64677-0\_2
- [17] Gyrard, A., Zimmermann, A., Sheth, A.: Building IoT-Based Applications for Smart Cities: How Can Ontology catalogs Help?. In: IEEE Internet of Things Journal 5(5), pp. 3978-3990. (2018). doi:10.1109/JIOT.2018.2854278
- [18] Minon, R., Paterno, F., Arrue, M.: An Environment for Designing and Sharing Adaption Rules for Accessible Applications. In: EICS'13. (2013)
- [19] Hubl, M.: Adaption rule for simultaneous use of smart urban objects from a fairness perspective. In: Proceedings of the 20th IEEE International Conference on Business Informatics (CBI 2018), pp. 89-98. (2018). doi:10.1109/CBI.2018.00019
- [20] Koch, M., Kötteritzsch, A., Fietkau, J.: Information radiators: using large screens and small devices to support awareness in urban space. In: Proceedings of the International Conference on Web Intelligence (WI '17), pp. 1080-1084 (2017). doi:10.1145/3106426.3109039

# An Adaptation of IoT to Improve Parcel Delivery System

Ha Yoon Song

Department of Computer Engineering,  
Hongik University, Seoul, Republic of Korea  
Email: hayoon@hongik.ac.kr

Hyochang Han

Department of Electronic and Electrical Engineering,  
Hongik University, Seoul, Republic of Korea  
Email: hcmiso1@gmail.com

**Abstract**—Recently, IoT technology has been applied in various field. One of the possible fields of an application is logistics system. In current system, a delivery must go through the designated logistics hub, which doesn't provide shortest distance. Such system costs time and inefficient expenses. In this paper, we propose an enhanced parcel delivery system based on IoT technology for reducing total delivery distance and seeking for much economy. First, we designed a sort of IoT devices which can be attached to parcels. This device has various functionalities including the ability to figure out current delivery route. Second, we addressed some difficulties such as : (i) issues linking IoT device into its platform; (ii) issues for designing IoT devices functionalities. Third, we propose ways to improve the efficiency of IoT based parcel delivery system. From these considerations, our system may improve total economics of parcel delivery system.

## I. INTRODUCTION

Thanks to the 4th Industrial Revolution, Internet of Things (IoT) technology is expanding and prospering in many industrial sectors. Accordingly, a rising number of domestic and foreign corporate are rushing for their own IoT platform in order to launch new services. In current IoT platform, computing and saving process of data is mostly done in central cloud. However, the centralized IoT system is causing many issues. Centralized IoT system requires a giant central server which processes and saves data that are received from a number of devices. This demands big administrative expense. It is difficult to increase connection of IoT devices continuously, since central server has its limit. To resolve this problem, we should expand the central server, but expanding the central server is very inefficient and will not be recommended. In addition, IoT devices require efficient management of data and central server is very important due to its real time data processing. Therefore, if there is a problem with the central server, every IoT device belongs to the platform will become useless. It is predicted that distributed networking of IoT will take place in the future to solve these problems [1].

In this case, the central server's role will be reduced, and the portion of the work handled by the terminal device will increase. In the future era of IoT, firmware-based IoT devices, which only deal with simple tasks, will be raised to the

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (NRF-2019R1F1A1056123).

level of the RTOS (Real Time Operation System) to secure safety and connectivity. Amazon acquired FreeRTOS, the most commonly used RTOS in embedded systems in 2017 [2]. In the same year, Google announced Tensorflow-Lite, a deep learning framework for IoT devices, to compensate for problems with cloud-based AI devices [3]. Based on existing experimental IoT platform provide by SK Telecom (SKT) of Republic of Korea, our parcel delivery enhancement can be implemented. The name of IoT platform is ThingPlug which also provides LoRaWAN as dedicated global network for IoT.

In this paper, we propose a distributed networked IoT delivery system instead of a centralized IoT platform as described in section II. Section III introduces functionalities of delivery dedicated IoT devices. Section IV introduces a process that connects IoT devices to distributed network IoT platform. Section V discusses issues in the process of interworking commercial networks. Section VI describes dedicated IoT devices we implemented, and Section VII concludes and introduces direction of research.

## II. IOT PLATFORM

Various factors such as Network, Device, and Application Server are needed to implement in IoT service. IoT platform refers to a service that allows various components to meet and easily combine and helps increasing utility value by connecting each element. As a representative, Qualcomm's AllJoyn, Microsoft's Azure IoT Suite and SKT's ThingPlug provide centralized services [4][5]. In this paper, we use SKT's ThingPlug as an IoT platform. Designed terminal device collects sensor data according to purpose and sends it to ThingPlug network server through gateway. Application Server can import data stored in ThingPlug's Network server and send control commands to the terminal device. These sequences are well expressed in Fig. 1

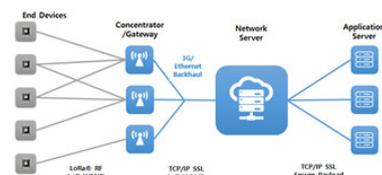


Fig. 1. IoT Platform

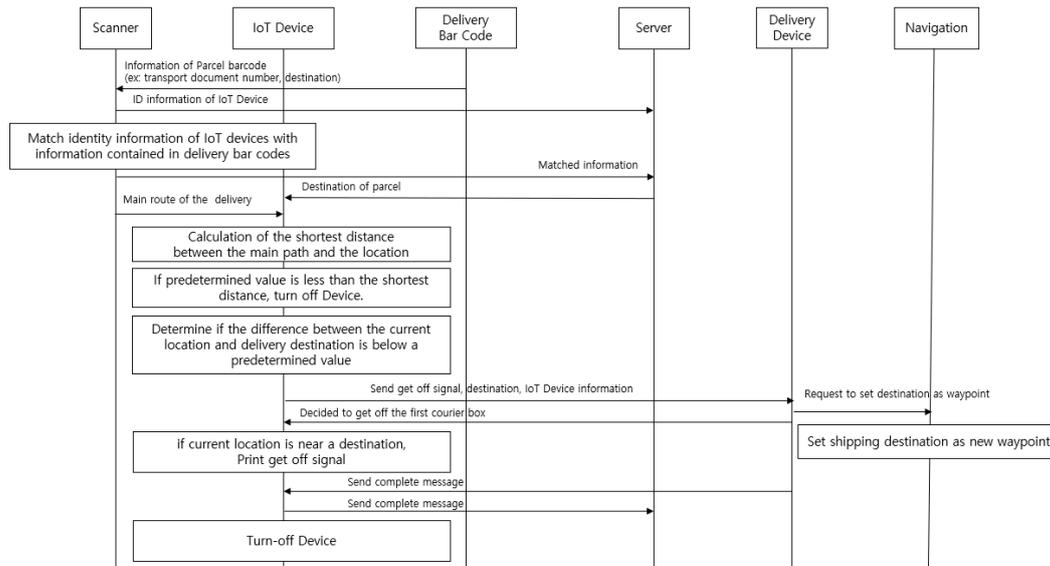


Fig. 2. Sequence Diagram of Improved Parcel Delivery

There are three ways to utilize ThingPlug Network Server. First case is to use one terminal device. In this case, we can communicate with ThingPlug Network Server using API provided or with ThingPlug via LoRa module. Second case is to use the terminal devices near the fixed gateway. In this case, we can connect the terminal device and the gateway in a wireless star topology method. Then the gateway collects the sensor from the terminal devices and send it to the ThingPlug server. At this time, the wireless communication between the terminal device and the Gateway can use LoRa RF, and the communication between the gateway and ThingPlug can use Ethernet. The third case is the terminal device that moves a wide range of areas. In this case, it is not possible for a single gateway to cover all the terminal devices, which is moving extensively. Therefore, we use Low Power Wide Area Network (LPWAN) that can communicate with ThingPlug server anywhere. LPWANs include LoRaWAN and Sigfox [6][7][8].

### III. IOT DEVICE FUNCTIONALITIES

Operation methods of IoT devices attached to delivery boxes were divided into several stages. There are four phases: the collection phase, GPS phase, the moving phase and the completion phase. We presented Sequence Diagram at Fig. 2

#### A. Parcel Collection Stage

ID information will be printed through the display module once pick-up of parcels starts with IoT device attached delivery box and IoT device is turned on. After a while LoRa will be connected with server. Then, the scanner will collect the data from the barcode attached on the delivery box. This data will be transmitted to the server and saved in a table form by matching the information, invoice number, and delivery destination of the IoT device. And IoT device get Collecting

path and calculate shortest path with collecting path and parcel destination. If that distance is less than threshold value, then start Parcel in move stage.

#### B. GPS Data Collection Stage

Once data are received from the geopositioning module of the IoT device, the data are stored for the current location of the IoT device. Usually, GPS is a representative one of geopositioning system. In case the data cannot be received from the GPS module, following steps are necessary. First, you need to get to nearby access point (AP) with Wi-Fi module. Then, obtain the current location information from the MAC address and API of the AP [9]. If Wi-Fi AP is not accessible, receive the current location information from Bluetooth module attached to the IoT device in the carrier's smartphone. If didn't get positioning data, you should be waiting for a certain delay and start from the beginning again.

#### C. Shortest Path Check Stage

In this stage, we need to know how to check the shortest path in order to add a new destination. At the destination, other parcels can just stop over and move on to next logistics hub. The process is as follows. Once data have been collected through stage B., then begin stage C. and use the data to check the distance between the current location and the destination. In case the distance is shorter than the threshold value, then send the parcel data to ThingPlug server via Lora and send ID info of IoT device to transmitter's smartphone via Bluetooth. Then, add new destination info from IoT device then update logistics path. Lastly, finish the moving step and move onto stage D. In case the distance is longer than the threshold value, loop will be started again after waiting for a certain delay of time.

#### D. Parcel Deliver Completion Stage

After the stage C. on every fixed time, then check the distance between current location and destination. In case the distance value is very small, IoT device will recognize that the parcel has arrived the destination. IoT device will emit signal by LED or display module. Then the courier can recognize that the box should be unloaded. Moreover, IoT device sends the completion signal to ThingPlug server and the signal will be updated. Then the server will send completion message to the IoT device. IoT device will emit return signal and will be turned off. Afterwards, the original shipper will collect IoT devices.

#### IV. IOT PLATFORM INTERLOCKING PROCESS

This paper adopted SK Telecom's IoT platform, ThingPlug, which utilizes LoRaWAN installed in South Korea, considering the size of the data and the size of the commercial network transmitted by IoT device. SK Telecom provides commercial network linkage through its officially certified LoRa module. In this paper, LoRa module of Wisol was used [10]. The LoRa module of the Wisol is connected to the Micro Controller Unit (MCU) and the Universal Asynchronous Receiver/Transmitter (UART) to send and receive messages through serial communication. The message sent from MCU to LoRa module follows the Command-Line Interface (CLI) command format defined in the user manual. ThingPlug's LoRa commercial network interworking is accomplished through Open Test Bed (OTB) certification and Quality Assurance (QA) testing.

Looking at the test items required by OTB certification, the first thing to do is to identify the debug message of the LoRa module from the MCU and design and implement it so that CLI command can be sent. During OTB authentication, all debug messages sent by LoRa module must be printed out because the debug message is verified through the customer's Host PC. Second, if you receive the Reset Downlink control command, we must reset the module after five seconds of delay. If IoT device that is on commercial network shows abnormal symptoms of operation, ThingPlug server sends an order to reset IoT device. MCU parses DEBUG message sent from LoRa module and performs Device Reset through CLI command. Since the reset command received from the server is a Confirmed message type, the reset must be performed after waiting about five seconds for LoRa module to receive the command and send the ACK. In addition, we should implement MCU command for perform the following actions: (1) Data Send 65 Bytes. (2)Data Send 66 Bytes. (3) Link Check Request. (4) Device Time Request.

For example, max payload is 65 bytes, so sending 65 bytes as shown in (1) can proceed without error but sending 66 bytes as shown in (2) must be able to check the ERROR debug message. Also, we need to check message type. There are two kind of message confirmed and unconfirmed. Confirmed message need to check packet was received. If the terminal device or Server sends a Confirmed message, they should send unconfirmed message with the ACK to make sure that message was processed. If there are no ACK, they send confirmed

message again. if retransmissions happened 8 times, return error and process will end. In addition, codes should be implemented that allow remote modification of the number of retransmissions. Because it is an additional remote-controlled IoT device, functions had to be implemented so that device can perform all tasks when variety of commands were issued from server. To do this, you need a code to verify that various functions work properly.

It was necessary to verify that Frame Count messages, which are larger than or smaller than the existing values, were received from the device in ThingPlug server by correctly parsing the corresponding messages, and that the frame Count was set to be larger than the existing values and then processed the message normally. It was also necessary to check if the message is handled normally when received from the terminal after setting up the Mgmt cmd. In addition, if the same message was received during processing after receiving the message, a code was required to confirm whether the message was dropped, and the uplink message was retransmitted and that the ACK was sent to the server.

There was a condition that the firmware of the LoRa module should be kept up to date to stabilize the IoT device, and that the PCB produced by itself should be used instead of the PCB provided with the IoT device. When communication with ThingPlug is required during the implementation of OTB certification items, open the device to the test network to conduct the experiment. The opening of the test network is completed by submitting the class with Device EUI to SK Telecom's ThingPlug manager and entering the information into LoRa module. You can register the device in your account in ThingPlug portal and send various downlink messages to the device through Open API Test. After OTB certification, SK Telecom is assigned a manager according to IoT service field and provided a test number for commercial network, which can be tested on commercial network in South Korea for three months after subscribing to CCBS. It also carries out commercialization and QA test through the manager's guide according to IoT service that it wants to provide. After passing QA test, connection of IoT device's commercial network will be completed.

#### V. ISSUES IN THE INTERWORKING PROCESS WITH COMMERCIAL NETWORK

In this paper, an esp32-based MCU board with built-in Wi-Fi function and Bluetooth function was designed and implemented. On top of this, GPS module and Wisol LoRa module for GPS tracking are connected by UART. However, there is only one pair of RX, TX pin for UART communication on pin diagrams. Therefore, the software serial replaces the deficient UART pins. Since esp32 does not officially provide the software serial, use the open source of third party. In addition, an additional hardware series is specified in the specification, which is not specified in the pin diagram. The esp32 offers up to three pairs of hardware serial ports.

During OTB test, the staff of the test shall be able to check all debug messages. Generally, the hardware serial buffer in

esp32 can be stored up to 256 bytes. If the LoRa module receives debug messages exceeding 256 bytes, the message is lost. Therefore, the size of the buffer should be increased to 1024 bytes through a member function of the Hardware Serial object to receive and output a complete debug message.

Since IoT devices are attached to delivery boxes, IoT devices must all be operational until delivery is completed after receiving delivery. To do so, battery management of IoT device is essential. For this purpose, the LoRa module OTB test-to-test Class A format was used. Class A format is not always on but is a very efficient method for battery management by storing data in the buffer when ACK comes from the server or when the value is entered from the sensor and performing operation every certain period of time. A typical method cannot store all data, but the same method of increasing the size of the hardware serial buffer in esp32 can store 1 kb of message, and the buffer of that capacity can store enough data, and the ACK of the server is sufficient. Energy management and performance are very important elements of IoT device, both aspects of IoT device performance and energy management can be grasped through this method.

## VI. IOT DEVICE

We implemented IoT Device for parcel delivery system as shown in Fig. 3. For basic control functions, MCU board with built-in esp32-based Wi-Fi and Bluetooth function, GPS module for GPS tracking function, and battery and battery charging are attached. The purpose of the design was to make the device as small as possible when manufacturing the device, and to prepare for the impact between delivery and delivery, the impact absorbing rubber was attached to the upper and lower parts of the IoT device, and acrylic plates and devices were firmly fixed using brass supports. To prevent impact damage to the acrylic or PCB plates, washers were added to each joint to enhance stability. In addition, the antenna was attached for desired Lora communication. The battery, which is at greater risk of short-circuit due to the weight during delivery, was fixed on acrylic plates. To secure GPS data, and the GPS Sensor was secured on acrylic plates to enable stable sensor reception. In order to manage the battery of the IoT device when not in use, the module can be switched on and off using the On - OFF switch. After checking whether the battery functions properly during delivery, we confirmed that 80 hours of continuous operation is practically possible. Although it has been suggested that the protruding part of the

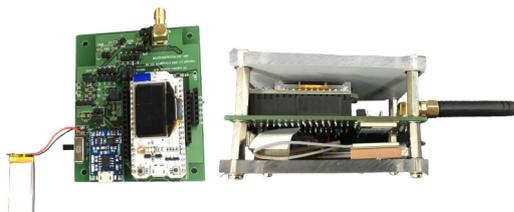


Fig. 3. Device Image

antenna will be a problem when using the IoT device seal, it is actually attached directly to the parcel when collecting data and can be used reliably when attaching it toward the sky when attached

## VII. CONCLUSION

In this paper, we proposed an enhanced parcel delivery system with IoT technology for reducing total delivery distance and for much economy. We showed the developing process and implementation to experimental IoT platform. Previously, central cloud did the most of control processing and computation. However, it is possible to divide up the processing to IoT device by using Bluetooth and Wi-Fi features. Then it became possible to process the data without going through server due to various functionalities of IoT device. In addition, the corresponding module can be deemed to be highly useful, such as by changing it to the delivery system or the tracking system. By using location information generated from these systems, it can be used as an active test data set in hub-related papers to prove the efficiency of the Hub and spoke method currently used and could check other additional logistics delivery methods [11][12]. Additionally, it would also devise and utilize a method for the security and integrity of the logistics system by combining blockchain with IoT platform and our IoT device functionality [13]. More functionalities can be considered; we can also add functions to identify the weight, type, size, and delivery area of the product entered before delivery and to show how to increase space efficiency and speed of delivery during loading parcels.

## REFERENCES

- [1] S. Tomovic, K. Yoshigoe, I. Maljevic, and I. Radusinovic, "Software-defined fog network architecture for iot," *Wireless Personal Communications*, vol. 92, no. 1, Jan 2017. doi: 10.1007/s11277-016-3845-0
- [2] R. Barry, "Freertos," <https://www.freertos.org/>, 2017.
- [3] Google, "Tensorflow-lite," <https://www.tensorflow.org/lite/guide>, 2018.
- [4] Microsoft, "Azure iot suite," <https://azure.microsoft.com>, 2018.
- [5] SKTelecom, "Thingplug," <https://sandbox.sktiot.com/>, 2018.
- [6] L. Alliance, "Lorawan spec 1.0.2," <https://lorawan-alliance.org/resource-hub/lorawan-specification-v102>, 2016.
- [7] A. Augustin, J. Yi, T. Clausen, and W. M. Townsley, "A study of lora: Long range and amp; low power networks for the internet of things," *Sensors*, vol. 16, no. 9, 2016. doi: 10.3390/s16091466
- [8] T. M. W. 1.0, "A technical overview of lora® and lorawan™," <https://www.everythingrf.com/whitepapers/details/2682-a-technical-overview-of-lora-and-lorawan>, 2015.
- [9] Sheng-Cheng Yeh, Wu-Hsiao Hsu, Ming-Yang Su, Ching-Hui Chen, and Ko-Hung Liu, "A study on outdoor positioning technology using gps and wifi networks," in *2009 International Conference on Networking, Sensing and Control*, March 2009. doi: 10.1109/ICNSC.2009.4919345
- [10] Wisol, "Lom102a user manual," <http://lora-support.wisol.co.kr/>, 2017.
- [11] G. F. George Deltas, Klaus Desmet, "Hub-and-spoke free trade areas: theory and evidence from israel," *Canadian Journal of Economics C.R.D.E., Universite de Montreal P.O. Box 6128, Station Centre-Ville Montreal, Quebec, H3C 3J7 Canada*, 13 August 2012. doi: <https://doi.org/10.1111/j.1540-5982.2012.01722.x>
- [12] G. G. Das and S. Andriamananjara, "Hub-and-spokes free trade agreements in the presence of technology spillovers: An application to the western hemisphere," *Review of World Economics*, Apr 2006. doi: <https://doi.org/10.1007/s10290-006-0056-x>
- [13] S. Huh, S. Cho, and S. Kim, "Managing iot devices using blockchain platform," in *2017 19th International Conference on Advanced Communication Technology (ICACT)*, Feb 2017. doi: <https://doi.org/10.23919/icact.2017.7890132>

# On Coverage of 3D Terrains by Wireless Sensor Networks

Mostefa Zafer

Ecole nationale Supérieure d'Informatique,  
BP 68M, 16309, Oued-Smar, Alger, Algérie,  
Email: m\_zafer@esi.dz

Mustapha Reda Senouci, Mohamed Aissani

Ecole Militaire Polytechnique,  
BP 17, 16046, Bordj El-Bahri, Alger, Algérie,  
Email: {mrsenouci, maissani}@gmail.com

**Abstract**—The coverage of a Region of Interest (RoI), that must be satisfied when deploying a Wireless Sensor Network (WSN), depends on several factors related not only to the sensor nodes (SNs) capabilities but also to the RoI topography. This latter has been omitted by most previous deployment approaches, which assume that the RoI is 2D. However, some recent WSNs deployment approaches dropped this unrealistic assumption. This paper surveys the different models adopted by the state-of-the-art deployment approaches. The weaknesses that need to be addressed are identified and some proposals expected to enhance the practicality of these models are discussed.

## I. INTRODUCTION

THE WSN deployment on 3D RoI presents many difficulties evoked by the topography, which has a direct impact on the coverage quality. Indeed, the presence of obstacles can hinder the detection of the target event. Also, the use of mobile SNs to eliminate the coverage voids becomes increasingly difficult, and the random deployment generates a very low coverage quality for 3D terrains [1]. Thus, ensuring the coverage of a 3D RoI by a WSN requires a deterministic deployment of the SNs, which consists in precomputing their number and their positions [2]. The resolution of this NP-hard problem [3] goes through a formulation phase, which describes the impact of the different factors on the coverage quality, and provides in its end an expression measuring the coverage quality produced by a given deployment scheme.

Once formulated, the problem is solved using heuristics or meta-heuristics, to select an appropriate deployment scheme. The practicality of a selected solution depends on the formulation process, which has been accomplished differently in the literature, depending on the factors taken into account and the modeling of their impacts on the coverage quality. When deploying WSNs on 3D RoIs, the formulation phase needs to be enriched by adding the RoI topography factor. To do this, it is necessary as a first step to model the SNs sensing capability, taking into account the RoI impact. In the second step, this model is used to formulate the RoI coverage. In this paper, we survey existing formulations of the coverage of 3D RoIs by WSNs in order to identify their main shortcomings that must be eliminated and make them more realistic. The coverage models and the RoI coverage deduction are detailed in Sections II and III, respectively. Section IV discusses the reliability of these models and some open issues. Section V concludes the paper.

## II. COVERAGE MODELS

To estimate the coverage quality of the RoI, produced by the deployment of a WSN composed of  $\mathcal{N}$  SNs, it is necessary to check the coverage status of each point  $p_i$  of the RoI. This status is deduced from a basic information  $\mathcal{C}(p_i, s_j)$ , which is the state of coverage of  $p_i$  by each SN  $s_j$ . The factors considered in estimating  $\mathcal{C}(p_i, s_j)$ , and the formulation of their impacts on  $\mathcal{C}(p_i, s_j)$  represent the coverage model [4]. In the existing coverage models,  $\mathcal{C}(p_i, s_j)$  is formulated according to one or more of the following factors: (i) the sensing range of  $s_j$ ; (ii) the sensing angle of  $s_j$ ; (iii) the topography of the RoI; (iv) the weather permeability of the RoI, and (v) the permeability of the objects separating  $p_i$  and  $s_j$ . It should be noted that the first and second factors are part of the SNs characteristics, while the third, fourth, and fifth factors belong to the RoI characteristics. Consequently, the most general formula of  $\mathcal{C}(p_i, s_j)$  is given by Eq. 1, where  $\mu_d(p_i, s_j)$ ,  $\mu_\phi(p_i, s_j)$ ,  $\mu_v(p_i, s_j)$ ,  $\mu_w(p_i, s_j)$ , and  $\mu_o(p_i, s_j)$  are binary or probabilistic functions, modeling the impact of the first, second, third, fourth, and fifth factors, respectively. In the sequel, we discuss the different models proposed to consider the impact of these factors on  $\mathcal{C}(p_i, s_j)$ .

$$\mathcal{C}(p_i, s_j) = \mu_d(p_i, s_j) \times \mu_\phi(p_i, s_j) \times \mu_v(p_i, s_j) \times \mu_w(p_i, s_j) \times \mu_o(p_i, s_j) \quad (1)$$

### A. Impact of the SNs sensing range

The sensing range of  $s_j$  is a reference distance  $r_s$  from which we can pronounce on the coverage of  $p_i$  by  $s_j$  in function of their distance  $d(p_i, s_j)$  [5], [6], [7], [8]. The influence of this factor on  $\mathcal{C}(p_i, s_j)$ , modeled by the function  $\mu_d(p_i, s_j)$ , takes three forms: (i) *Deterministic impact* [5], [6], where  $\mathcal{C}(p_i, s_j)$  is constant with respect to  $d(p_i, s_j)$ , as long as  $p_i$  is in the sensing range of  $s_j$ . Otherwise,  $\mathcal{C}(p_i, s_j)$  is null; (ii) *Probabilistic impact* [9], [10], [7], where  $\mathcal{C}(p_i, s_j)$  degrades with respect to  $d(p_i, s_j)$ , and it becomes null when the point  $p_i$  is outside the sensing range of  $s_j$ ; (iii) *Hybrid impact* [11], [12], [13], by considering that  $s_j$  has two sensing ranges, the first is "with certitude", noted  $r_1$ , and the second is "without certitude", noted  $r_2$ , where  $r_2 > r_1$ . Thus,  $\mathcal{C}(p_i, s_j)$  is constant with respect to  $d(p_i, s_j)$ , as long as  $p_i$  is in the sensing range "with certitude" of  $s_j$ ; it is null when  $p_i$  is outside the sensing range "without certitude" of  $s_j$ , and it degrades with

respect to  $d(p_i, s_j)$ , in the remaining case. Fig. 1 shows 2D-graphical representations of these models.

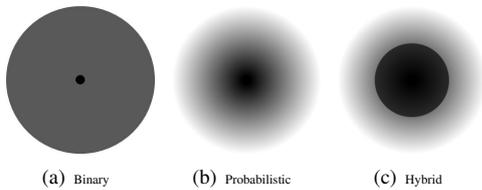


Fig. 1. Impact of the sensing range.

### B. Impact of the SNs sensing angle

Sometimes, the sensing capability of SNs is limited to an angle  $\phi$ . In this case, the coverage of  $p_i$  by  $s_j$  depends on the position of  $p_i$  with respect to the orientation of  $s_j$ , represented by the unit vector  $\vec{u}_s$ . Thus, the position of  $p_i$  with respect to the orientation of  $s_j$ , is measured by the angle  $(\vec{u}_s, \vec{s}_j \vec{p}_i)$ , whose impact on  $\mathcal{C}(p_i, s_j)$  takes three forms: (i) *Deterministic impact* [14], [6], [15], which means that  $\mathcal{C}(p_i, s_j)$  is constant with respect to  $(\vec{u}_s, \vec{s}_j \vec{p}_i)$ , as long as  $p_i$  is in the sensing angle of  $s_j$ . Otherwise,  $\mathcal{C}(p_i, s_j)$  is null; (ii) *Probabilistic impact* [9], [10], which means that  $\mathcal{C}(p_i, s_j)$  degrades in function of  $(\vec{u}_s, \vec{s}_j \vec{p}_i)$ , and it becomes null when  $p_i$  is outside the sensing angle of  $s_j$ ; (iii) *Hybrid impact* [16], [17], which means that  $s_j$  has two sensing angles, the first is “with certitude”, noted  $\phi_1$ , and the second is “without certitude”, noted  $\phi_2$ , where  $\phi_2 > \phi_1$ . Thus,  $\mathcal{C}(p_i, s_j)$  is constant with respect to  $(\vec{u}_s, \vec{s}_j \vec{p}_i)$ , as long as  $p_i$  is in the sensing angle “with certitude” of  $s_j$ ; it is null when  $p_i$  is not in the sensing angle “without certitude” of  $s_j$ ; and it degrades depending on  $(\vec{u}_s, \vec{s}_j \vec{p}_i)$ , in the remaining case. Fig. 2 shows 2D-graphical representations of these models.

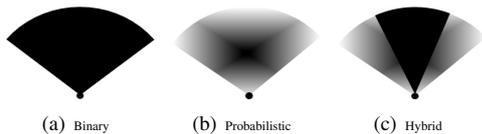


Fig. 2. Impact of the sensing angle.

### C. Impact of the RoI topography

To take into account the topography impact on  $\mathcal{C}(p_i, s_j)$ , most formulations proceed to verify the “visibility” between  $p_i$  and  $s_j$ , by using the Line of Sight (LoS) [15], [18], [19] method (Fig. 3). This latter selects  $N_v$  points  $q_i(x_i, y_i, z_i)_{1 \leq i \leq N_v}$ , located on the segment  $[p_i, s_j]$ , and compares the altitude  $z_i$  of each point  $q_i$  with the RoI height  $\mathcal{E}(x_i, y_i)$ , provided by a terrain model  $\mathcal{E}$ . Thus,  $p_i$  and  $s_j$  are considered inter-visible, if each point  $q_i$  is above the terrain. Once the visibility between  $p_i$  and  $s_j$  is verified, its impact on  $\mathcal{C}(p_i, s_j)$  takes two forms: (i) *Deterministic impact* [20], [15], [18], [19], by considering that  $p_i$  can be covered by  $s_j$ , only if  $p_i$  and  $s_j$  are inter-visible. In the opposite case,

$\mathcal{C}(p_i, s_j)$  is null. (ii) *Probabilistic impact* [14] by considering that  $\mathcal{C}(p_i, s_j)$  deteriorates (does not cancel out) according to the number of obstacles separating  $p_i$  and  $s_j$ . Some WSNs deployment approaches on 3D terrains do not consider the visibility factor when formulating the coverage. This choice is based on one of the following justifications. (i) Some events are detectable even if they occur in locations invisible to the SNs [21]; (ii) The terrain is assumed to be sufficiently convex, so that the visibility between a SN and any point within its sensing range is always possible [3], [1], [22]; (iii) The impact of the RoI topography, is already taken into account during the parameterization of  $\mathcal{C}(p_i, s_j)$  depending on  $d(p_i, s_j)$  [6], [7].

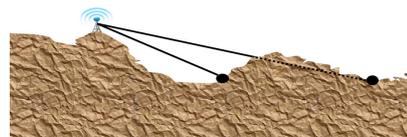


Fig. 3. LoS concept.

### D. Impact of the RoI weather and the RoI objects

In [23],  $\mathcal{C}(p_i, s_j)$  is considered dependent on the permeability of the objects and the weather of the RoI, where their impacts on  $\mathcal{C}(p_i, s_j)$  is formulated separately in a probabilistic way.

On Table I, we list the different coverage models, denoted  $C_1, C_2, \dots, C_{12}$ , adopted by the WSNs deployment approaches on 3D surfaces, where the difference between them lies in the factors taken into account and how their influences on  $\mathcal{C}(p_i, s_j)$  is modeled. The type of each model, which can be binary or probabilistic, is determined by the possible values of  $\mathcal{C}(p_i, s_j)$ . Indeed, a coverage model is binary if  $\mathcal{C}(p_i, s_j) \in \{0, 1\}$ , and it is considered probabilistic if  $\mathcal{C}(p_i, s_j) \in [0, 1]$ .

## III. ROI COVERAGE

The basic information  $\mathcal{C}(p_i, s_j)$  is used to formulate the coverage state  $Cov(p_i, \mathcal{N})$  of each point  $p_i$  with respect to the WSN composed of  $\mathcal{N}$  SNs. If the adopted coverage model is binary [6],  $p_i$  can be in two states with respect to the WSN: “ $p_i$  is covered”, if it is covered by at least one SN, and “ $p_i$  is not covered”, if it is not covered by any SN. Therefore,  $Cov(p_i, \mathcal{N}) = \max_{1 \leq j \leq \mathcal{N}} \mathcal{C}(p_i, s_j)$ . If the adopted coverage model is probabilistic, and  $\mathcal{C}(p_i, s_j)$  is interpreted as the probability of coverage of  $p_i$  by  $s_j$  [14],  $Cov(p_i, \mathcal{N})$  is given by  $Cov(p_i, \mathcal{N}) = 1 - \prod_{1 \leq j \leq \mathcal{N}} (1 - \mathcal{C}(p_i, s_j))$ . If the adopted coverage model is probabilistic and  $\mathcal{C}(p_i, s_j)$  is interpreted as the coverage quality of  $p_i$  by  $s_j$  [23], [11], [12],  $Cov(p_i, \mathcal{N})$  is equal to  $\max_{1 \leq j \leq \mathcal{N}} \mathcal{C}(p_i, s_j)$ , which means that the coverage of  $p_i$  is assigned to the SN that offers the best coverage quality for  $p_i$ . The last step in the formulating phase is to express the coverage quality  $Cov(\mathcal{A}, \mathcal{N})$  of the RoI  $\mathcal{A}$  by the WSN composed of  $\mathcal{N}$  SNs, using the state  $Cov(p_i, \mathcal{N})$  of each point  $p_i \in \mathcal{A}$ . This step is strongly related to the terrain model  $\mathcal{E}$  adopted to represent  $\mathcal{A}$  (Fig. 4), which may

TABLE I  
 VARIOUS COVERAGE MODELS USED IN THE LITERATURE.

Model	Type	Impact of considered factors					References
		SNs characteristics		RoI characteristics			
		Detection range	Detection angle	Topography	Weather	Object	
$C_1$	Binary	Binary	.	.	.	[3], [1], [5], [8], [24], [25]	
$C_2$	Binary	Binary	.	Binary	.	[26], [20], [18], [19], [4], [27]	
$C_3$	Probabilistic	Probabilistic	.	.	.	[22], [7], [28]	
$C_4$	Probabilistic	Hybrid	.	.	.	[21]	
$C_5$	Probabilistic	Hybrid	.	Binary	.	[11], [12], [13], [29]	
$C_6$	Probabilistic	Hybrid	Hybrid	Binary	.	[16], [17], [30]	
$C_7$	Binary	Binary	Binary	.	.	[6]	
$C_8$	Binary	Binary	Binary	Binary	.	[15], [31]	
$C_9$	Probabilistic	Binary	Binary	Probabilistic	.	[14]	
$C_{10}$	Probabilistic	Probabilistic	Probabilistic	Binary	.	[9], [10]	
$C_{11}$	Probabilistic	Binary	Binary	Binary	Probabilistic	Probabilistic	[23]
$C_{12}$	Probabilistic	Hybrid	Probabilistic	Binary	.	[32]	

be discontinuous, such as the matrix model or continuous, such as the mathematical and the TIN (Triangulated Irregular Network) models.

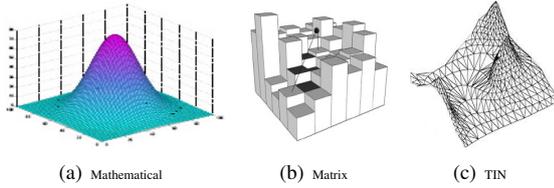


Fig. 4. Various terrain models used in the literature.

In the matrix model,  $Cov(\mathcal{A}, \mathcal{N})$  represents the rate of points of  $\mathcal{E}$  covered by the  $\mathcal{N}$  SNs [11], [12], [15], [18], considering that these points have the same level [11], [12], [13], [15], [18] or different levels [10] of importance. Thus,  $Cov(\mathcal{A}, \mathcal{N})$  is given by Eq. 2, where the weight  $w_i$  represents the importance assigned to a point  $p_i \in \mathcal{E}$ .

$$Cov(\mathcal{A}, \mathcal{N}) = \frac{\sum_{p_i \in \mathcal{E}} w_i \cdot Cov(p_i, \mathcal{N})}{\sum_{p_i \in \mathcal{E}} w_i} \quad (2)$$

Most approaches based on the mathematical model [5], [28] construct a matrix model  $\hat{\mathcal{E}}$  from the original model  $\mathcal{E}$ . Hence,  $Cov(\mathcal{A}, \mathcal{N})$  is computed according to Eq. 2, considering only the points of  $\hat{\mathcal{E}}$ . This transformation is avoided in some approaches [6]. The alternative idea is to estimate, using a geometric calculation, the surface  $\|\mathcal{A}_j\|$  of the portion  $\mathcal{A}_j \subset \mathcal{A}$  covered by each SN  $s_j$ . Thus,  $Cov(\mathcal{A}, \mathcal{N})$  is computed as the ratio of the surface covered by the SNs and the surface  $\|\mathcal{A}\|$  of  $\mathcal{A}$  (Eq. 3).

$$Cov(\mathcal{A}, \mathcal{N}) = \frac{\|\bigcup_{1 \leq j \leq \mathcal{N}} \mathcal{A}_j\|}{\|\mathcal{A}\|} \quad (3)$$

The approaches based on the TIN model formulate firstly the coverage quality  $Cov(t_i, \mathcal{N})$  of each triangle  $t_i \in \mathcal{E}$  [3], [1], [22], [23], which is calculated as the average of the coverage

quality of the important points of  $t_i$  (its center of gravity and its vertexes). After that,  $Cov(\mathcal{A}, \mathcal{N})$  is computed as the average of the coverage qualities of all the triangles  $t_i \in \mathcal{E}$ , assuming that these triangles have the same level of importance [4] or different levels [23]. The formula of  $Cov(\mathcal{A}, \mathcal{N})$  is given by Eq. 4, where  $w_i$  represent the importance affected to the triangle  $t_i \in \mathcal{E}$ .

$$Cov(\mathcal{A}, \mathcal{N}) = \frac{\sum_{t_i \in \mathcal{E}} w_i \cdot Cov(t_i, \mathcal{N})}{\sum_{t_i \in \mathcal{E}} w_i} \quad (4)$$

#### IV. DISCUSSIONS

The first important remark concerns the choice of the terrain model to represent the RoI. This choice is based, in most cases, on criteria other than reliability. For instance, the matrix model was used in [14], [9], although the TIN model is more appropriate. Additionally, the assumptions adopted in some formulations [3], [1] to justify the omission of the visibility factor are unfounded. Indeed, the assumption that the RoI is sufficiently smooth, where the visibility between a SN and all the points in its sensing range is guaranteed, is not realistic. Moreover, in the case where the visibility factor is considered [23], [14], its impact is formulated independently of the nature of the phenomenon being monitored and the type of SNs (laser, radio, etc.) used to detect it. In fact, these two parameters determine the manner (probabilistic or deterministic) and the degree of visibility impact on the coverage quality.

Moreover, some formulations adopt clearly unrealistic assumptions, such as the deterministic impact of the various factors on the coverage quality [3], [1], as well as the omnidirectional sensing capability of the SNs [22], [21]. Furthermore, most of the formulations [5], [6] do not consider the constraints imposed by the RoI, which limits the possible positions of the SNs. The existence of such constraints in complex 3D terrains is very likely. As a result, we believe that the design of a realistic coverage model remains an open issue. Its resolution requires to consider not only the type of SNs to be used but also the phenomenon to be monitored as well

as experimental tests, allowing the correct parameterization of their impacts and the deduction of the real relationship between them.

## V. CONCLUSION

Examining the several formulations of 3D terrain coverage by WSNs allowed us to confirm that this paramount process remains an open issue. We believe that carrying out experimental tests, to correctly assess and model both the influence of the above-discussed factors and the real relationship between them is the first step towards a practical and effective solution. As a future work, we plan to carry out an in-depth evaluation of existing resolution approaches related to the problem of 3D terrains coverage by WSNs. This will allow us to gain a better understanding in order to provide additional guidelines on the appropriate choice of the modeling/resolution approaches.

## REFERENCES

- [1] L. Kong, M.-C. Zhao, X.-Y. Liu, J. Lu, Y. Liu, M.-Y. Wu, and W. Shu, "Surface Coverage in Sensor Networks," *TPDS*, vol. 25, no. 1, 2014. doi: 10.1109/TPDS.2013.35
- [2] M. R. Senouci and A. Mellouk, *Deploying Wireless Sensor Networks: Theory and Practice*. Elsevier, 2016. ISBN 978-1-78548-099-7
- [3] M.-C. Zhao, J. Lei, M.-Y. Wu, Y. Liu, and W. Shu, "Surface Coverage in Wireless Sensor Networks," in *INFOCOM*, 2009. doi: 10.1109/INFCOM.2009.5061912
- [4] M. Zafer, M. R. Senouci, and M. Aissani, "Terrain Partitioning Based Approach for Realistic Deployment of Wireless Sensor Networks," in *CIIA'18*, 2018. doi: 10.1007/978-3-319-89743-13\_7
- [5] K. Kim, "Mountainous terrain coverage in mobile sensor networks," *IET Comm.*, vol. 9, no. 5, 2015. doi: 10.1049/iet-com.2014.0443
- [6] F. Xiao, X. Yang, M. Yang, L. Sun, R. Wang, and P. Yang, "Surface Coverage Algorithm in Directional Sensor networks for Three-Dimensional Complex Terrains," *TST*, vol. 21, no. 4, 2016. doi: 10.1109/TST.2016.7536717
- [7] T. Song, C. Gong, and C. Liu, "A practical coverage algorithm for wireless sensor networks in real terrain surface," *IJWMC*, vol. 5, no. 4, 2012. doi: 10.1504/IJWMC.2012.051514
- [8] F. Li, J. Luo, W. Wang, and Y. He, "Autonomous Deployment for Load Balancing k-Surface Coverage in Sensor Networks," *TWC*, vol. 14, no. 1, 2015. doi: 10.1109/TWC.2014.2341585
- [9] V. Akbarzadeh, C. Gagné, M. Parizeau, and M. A. Mostafavi, "Black-box Optimization of Sensor Placement with Elevation Maps and Probabilistic Sensing Models," in *Int. Symp.*, 2011. doi: 10.1109/ROSE.2011.6058544
- [10] V. Akbarzadeh, C. Gagné, M. Parizeau, M. Argany, and M. A. Mostafavi, "Probabilistic Sensing Model for Sensor Placement Optimization Based on Line-of-Sight Coverage," *ToIM*, vol. 62, no. 2, 2013. doi: 10.1109/TIM.2012.2214952
- [11] N. Unaldi, S. Temel, and V. K. Asari, "Method for Optimal Sensor Deployment on 3D Terrains Utilizing a Steady State Genetic Algorithm with a Guided Walk Mutation Operator Based on the Wavelet Transform," *Sensors*, vol. 12, no. 4, 2012. doi: 10.3390/s120405116
- [12] S. Temel, N. Unaldi, and O. Kaynak, "On Deployment of Wireless Sensors on 3D Terrains to Maximize Sensing Coverage by Utilizing Cat Swarm Optimization with Wavelet Transform," *TSMC*, vol. 44, no. 1, 2014. doi: 10.1109/TSMCC.2013.2258336
- [13] Y. Hang, L. Xunbo, W. Zhenlin, Y. Wenjie, and H. Bo, "A Novel Sensor Deployment Method Based on Image Processing and Wavelet Transform to Optimize the Surface Coverage in WSNs," *CJE*, vol. 25, no. 3, 2016. doi: 10.1049/cje.2016.05.015
- [14] N. T. Tam, H. D. Thanh, L. H. Son, and V. T. Le, "Optimization for the sensor placement problem in 3D environments," in *ICNSC*, 2015. doi: 10.1109/ICNSC.2015.7116057
- [15] V. Akbarzadeh, A. H.-R. Ko, C. Gagné, and M. Parizeau, "Topography-Aware Sensor Deployment Optimization with CMA-ES," in *ICPPSN*, 2010. doi: 10.1007/978-3-642-15871-1\_15
- [16] B. Cao, J. Zhao, Z. Lv, and X. Liu, "3D Terrain Multiobjective Deployment Optimization of Heterogeneous Directional Sensor Networks in Security Monitoring," *TBD*, vol. 14, no. 8, 2015. doi: 10.1109/TB-DATA.2017.2685581
- [17] B. Cao, J. Zhao, Z. Lv, X. Liu, X. Kang, and S. Yang, "Deployment Optimization for 3D Industrial Wireless Sensor Networks Based on Particle Swarm Optimizers with Distributed Parallelism," *JNCA*, 2017. doi: 10.1016/j.jnca.2017.08.009
- [18] S. Doodmana, A. Afghantoloe, M. A. Mostafavi, and F. Karimipour, "3D extension of the VOR algorithm to determine and optimize the coverage of geosensor networks," in *ISPRS*, 2014. doi: 10.5194/isprsarchives-XL-2-W3-103-2014
- [19] A. H.-R. Ko and F. Gagnon, "Process of 3D wireless decentralized sensor deployment using parsing crossover scheme," *EACI*, vol. 11, 2015. doi: 10.1016/j.aci.2014.11.001
- [20] K. Veenstra and K. Obraczka, "Guiding Sensor Node Deployment Over 2.5D Terrain," in *ICC*, 2015. doi: 10.1109/ICC.2015.7249396
- [21] J.-H. Seo, Y. Yoon, and Y.-H. Kim, "An Efficient Large-Scale Sensor Deployment Using a Parallel Genetic Algorithm Based on CUDA," *IJDSN*, vol. 2016, 2015. doi: 10.1155/2016/8612128
- [22] M. Jin, G. Rong, H. Wu, L. Shuai, and X. Guo, "Optimal Surface Deployment Problem in Wireless Sensor Networks," in *INFOCOM*, 2012. doi: 10.1109/INFCOM.2012.6195622
- [23] H. R. Topcuoglu, M. Ermis, and M. Sifyan, "Positioning and Utilizing Sensors on a 3D Terrain Part I: Theory and Modeling," *TSMC*, vol. 41, no. 3, 2011. doi: 10.1109/TSMCC.2010.2055850
- [24] N. Boufares, I. Khoufi, P. Minet, and L. Saidane, "Covering a 3D flat surface with autonomous and mobile wireless sensor nodes," in *PEMWN*, 2017. doi: 10.1109/IWCMC.2017.7986528
- [25] C. Wang and H. Jiang, "SURF: A Connectivity-based Space Filling Curve Construction Algorithm in High Genus 3D Surface WSNs," in *CCC*, 2015. doi: 10.1109/INFCOM.2015.7218470
- [26] A. T. Murray, K. Kim, J. W. Davis, R. Machiraju, and R. Parent, "Coverage optimization to support security monitoring," *CEUS*, vol. 31, 2007. doi: 10.1016/j.compenvurbsys.2006.06.002
- [27] B. Cao, J. Zhao, P. Yang, Z. Lv, X. Liu, X. Kang, S. Yang, K. Kang, and A. Anvari-Moghaddam, "Distributed parallel cooperative coevolutionary multi-objective large-scale immune algorithm for deployment of wireless sensor networks," *FGCS*, vol. 82, 2018. doi: 10.1016/j.future.2017.10.015
- [28] L. Feng, Z. Sun, and T. Qiu, "Genetic Algorithm-Based 3D Coverage Research in Wireless Sensor Networks," in *ICCISIS*, 2013. doi: 10.1109/CISIS.2013.112
- [29] N. Unaldi and S. Temel, "Wireless Sensor Deployment Method on 3D Environments to Maximize Quality of Coverage and Quality of Network Connectivity," in *WCECS*, 2014, pp. 1-6.
- [30] B. Cao, X. Kang, J. Zhao, P. Yang, Z. Lv, and X. Liu, "Differential Evolution-based 3D Directional Wireless Sensor Network Deployment Optimization," *JIoT*, vol. 5, no. 5, 2018. doi: 10.1109/JIoT.2018.2801623
- [31] A. Afghantoloe, S. Doodman, F. Karimipour, and M. A. Mostafavi, "Coverage Estimation of Geo-sensors in 3D Vector Environments," in *GIRC*, 2014. doi: 10.13140/2.1.2229.0723
- [32] M. Argany, F. Karimipour, F. Mafi, and A. Afghantoloe, "Optimization of Wireless Sensor Networks Deployment Based on Probabilistic Sensing Models in a Complex Environment," *JASN*, vol. 20, no. 7, 2018. doi: 10.3390/jasn7020020

# Smart Urban Objects to Enhance Safe Participation in Major Events for the Elderly

Tobias Zimpel, Marvin Hubl  
University of Hohenheim  
Stuttgart, Germany

Email: {tobias.zimpel, marvin.hubl}@uni-hohenheim.de

**Abstract**—IoT increasingly permeates the public area, e.g., in traffic control and public transport. We propose to equip conventional urban objects with IoT technology to transform them into *Smart Urban Objects (SUO's)*. While there exists some research exploring the potentials, specific solutions to enhance safety for the elderly outdoors are still lacking. The elderly's safety is threatened due to declining physical conditions. As a consequence, the elderly may be excluded from outdoor activities such as participating in major events. Against this backdrop, we design SUOs for adaptive indications of urban hazards, barrier-free passages and for smart reservation of seats to enhance resting possibilities. We report on our solution using Bluetooth technology for remote sensing of older pedestrians serving as input for the objects' adaptive capacities. The SUOs have been installed for test purposes on a major event in a larger German city.

## I. INTRODUCTION

GROWING older is—sooner or later—inevitably accompanied by a deterioration of life skills, concerning motor skills, information processing skills and sensory capabilities [1]. This regularly intensifies the individual perception of threats to safety, particularly outside of the own home. Changes in body mechanics and impaired endurance pose a substantial risk for safe mobility [2], [3]. As a consequence, the elderly tend to avoid going outdoors without active assistance. This may lead to declining cultural and social participation up until the feeling of isolation. Empirical research has broadly studied and confirmed the positive influence of outdoor activities for the elderly's well-being [4], [5], [6] and we suggest that IT use of the elderly can have a positive impact on their participation in outdoor activities [7].

Demographic projections foresee a disparity between younger people who can provide care, and older people who will potentially be in need for care [8]. This anticipation virtually reinforces the requirement to find innovative means for assisting the elderly in their outdoor activities up until old age. The role of the built environment in this respect has long been acknowledged [9], [10]. As particular assistive means in the built environment we design so called Smart Urban Objects.

*Smart Urban Objects (SUO's)* are urban objects equipped with sensors, actuators and enabled to make potentially use of digital information processing. Examples of such SUOs are “smart” park benches, street lights, information panels or parking lots [11], [12], [13], [14]. By interconnecting them via internet technology they are *IoT objects*.

Unlike the example SUOs, we specifically aim at enhancing safety of older pedestrians for the participation in public outdoor major events. These events can be particularly exhaustive and hence hazardous for the elderly because of several reasons:

On major events use to be a relatively large crowd of people on a limited space. Additionally, the event area is usually characterized by temporary installations. These installations are furthermore rarely tailored for accessibility. All this may lead to increased confusion respectively to increased mental exhaustion and to increased physical exhaustion.

The arrangement of temporary objects also produces unexpected risks of tripping, as for example with cable bridges. Because of the temporary character, people often do not exactly know, where they need to go, if they search something particular. However, even if people know where they need to go, they are often unaware about accessible paths on the temporarily arranged area.

Older pedestrians may have increased need for seating rests. One reason is the aforementioned mental and physical exhaustion but can also be due to dizziness. Weather, especially heat, may be a factor, too. Older people often have impaired thermoregulation because of reduced fluid balance [15], [16]. This may stray the circulatory system even stronger, leading to seating rests for its relief and recovery.

Therefore, we design SUOs that adaptively warn for potential tripping hazards, indicate accessible paths and allow for reservation of seats from anywhere on the area. We have prototypically tested the SUOs on an outdoor public major event in a larger German city.

We make use of Leveson's conceptualization for *safety engineering* [17]. Safety is constituted as avoidance of accidents where the concept of accidents encompasses all situations that involve some unacceptable loss [17, p. 181]. In this respect, we focus on *conditions* potentially leading to accidents. Therefore, the concept of hazards is pivotal and is defined as “[a] system state or set of conditions that, together with a particular set of worst-case environmental conditions, will lead to an accident (loss)”. [17, p. 184]

On this basis, we make use of the implication that accidents occur only if an hazardous state coincides with some worst-case environmental conditions. Note, that hazards and environment constitute a dualism. That means whether a state is hazardous and whether an environmental condition is a worst-case one is mutually dependent. As our design object is the

urban *environment* through the means of SUOs, we define older pedestrians as the “systems” which can potentially be in a hazardous state.

We assume older pedestrians to be in a hazardous state if (a) they are only able to lift their legs comparatively little, (b) have impaired eyesight or (c) are in need for a rest. Corresponding worst-case conditions are (a) structural barriers, (b) “hidden” stumbling blocks and (c) missing seat opportunity. (a) Structural barriers, like curb stones or steps, may lead to an accident, if an older pedestrian cannot lift her legs high enough. (b) Stumbling blocks may lead to an accident if an older pedestrian cannot recognize it visually. (c) A need for a seat rest may lead to an accident, if there is no seat available.

In turn, (a) inability to lift the legs high is no problem if there are no steps or the like, (b) visual impairments do not lead to accidents if there is no hidden stumbling block on an older pedestrian’s path and (c) need for a seat rest is not critical if there is an available seat possibility. Conversely, if (a) a pedestrian can lift its legs high then steps are no effective barriers, if (b) stumbling blocks are visually recognized, they pose only little risks and if (c) a pedestrian has good endurance, missing seats are no safety problem.

This shows that principally two options are possible for enhancing safety: Avoiding hazardous states or avoiding worst-case environmental conditions. Since in our conception the hazardous states are inherent to the older pedestrians, we take them as given and seek to avoid the corresponding worst-case environmental conditions.

(a) The disability to lift the legs appropriately must be taken as is. However, the corresponding worst-case environmental condition can be avoided by guiding the pedestrians through passages without steps, curb stones and the like. Then, on the pedestrian’s individual path, there is no worst-case environmental condition. (b) Impaired vision must be taken as is. However, the corresponding worst-case environmental condition can be avoided by clearly indicating stumbling blocks, thus “unhide” them. There, we expect that an adaptive indicator is more salient than a static indicator. (c) Need for a rest, eventually, is taken as given. However, by reserving some seats and making them available adaptively to older pedestrians who announce a need for a rest, the worst-case condition that there is no available seat can be avoided.

Against this background we formulate our design-oriented research question:

*How to design Smart Urban Objects (SUO’s) for major events to adaptively avoid worst-case environmental conditions for older pedestrians being in a defined hazardous state?*

This paper proceeds as follows: In section II, we review the state of the art on IoT conceptualization as well as on pedestrian support with smart objects. In section III, we report on the design of our SUOs. In section IV, we evaluated our SUOs in terms of its principal functionality. In section V we discuss our SUOs and revised some conceptual issues,

like IoT. In section VI we conclude our work and provide an outlook.

## II. STATE OF THE ART

### A. *Internet of Things and Smart Objects*

Most basically “IoT is the network of things, with device identification, embedded intelligence, and sensing and acting capabilities, connecting people and things over the Internet.” [18, p. 4] As IoT objects are characterized by their situatedness in a physical real-world environment, sensors provide an input interfaces from the environment to the IoT object and actuators provide an output interface from the IoT object to the environment. Sensors convert physical signals from the real-world environment into digital data. Actuators convert digital data into actions that shall affect the environment.

In our conception, we allow that sensory input can be digitally pre-coded data, as for example when sensing radio signals, such as RFID or Bluetooth signals. While actuators are often considered to be physically moving parts [18, p.71], we include in our conception also non-moving parts that shall exert influence on the environment, like lights, audio-output, displays.

IoT can be seen from three perspective [19]: (a) The “Things”-oriented view contains technologies such as RFID, UID, wireless sensors and actuators as well as that the things shall be able to communicate with each other [19]. (b) The “Internet”-oriented view contains technologies such as IP for Smart Objects (IPSO) or Web of Things as well as middleware [19]. (c) Additionally a “Semantic”-oriented view can be taken, containing semantic technologies, e.g. for reasoning over data [19].

IoT conventionally has four layers [20]: (1) The sensor/actuator layer relates directly to the sensors, actuators as well as to the IoT objects themselves, hence to hardware [20]. (2) The network layer relates to the basic network technologies for data transfer [20]. (3) The interface layer provides methods for interactions with the IoT objects for other applications and users [20]. (4) Finally, the service provides services to satisfy user requirements [20]. This means, applications can be abstracted from the hardware-oriented sensor/actuator layer and be implemented on the service layer.

IoT objects are often referred to as smart objects. There are several differentiations for the “smartness” of the objects. One attempt is to differentiate the awareness capabilities [21]: (1) Activity-aware objects understand the environment as events that are directly linked with the object, such as touching the object [21]. (2) Policy-aware objects relate events to organizational policies. (3) Process-aware objects relate events to organisational processes [21]. Although this differentiation seems to focus on applications for business operations, it is conceptually applicable to other applications, e.g. if organizational policies are substituted by other norms.

An alternative typology of smart objects is given by their (I) capacity to store relevant data, including an identifier for themselves, (S) capabilities for sensory perceptions and (A) execution of actions with actuators, (D) decision-making

ability and finally (N) network connectivity [22]. Referring to the letters in brackets, an “I-N object” for example has an identity and data storage (where at least its identity is stored) and network connectivity. Note that not all combinations are considered to be realistic as for instance most object types without “I” [22]. Note also that objects need to implement a form of advanced information processing to exhibit decision-making abilities.

Smart objects are an important information technological basis for Smart City [23] as an application of IoT [24], [25]. Most conceptualization of Smart City contain or even accentuate inclusiveness as an aim. In this respect, we consider our IoT application to be a Smart City use case.

### B. Smart Urban Objects and Pedestrian Support

Poulsen et al. propose an urban light system that can respond adaptively to pedestrian’s occupancy patterns, wind velocity or that can be customized to individual color preferences via smart phone [12]. Albeit not in an outdoor setting, the potential effect of adaptive colors on the mood of seniors has been studied by Huldtgren et al. [26]. Cunha & Fuks propose to use light systems as a “host” for sensors to support continuous proactive care within a feedback loop [27].

Another type of objects are public interactive screens. These can be utilized as adaptive urban information panels. Cremonesi et al., Müller et al. as well as Vogel & Balakrishnan examine concepts for personalized interactions on such public screens [13], [28], [29]. The basic approach is to define virtual fences around the screen and for identified pedestrians in near proximity sections on the screen can be personalized.

As an interesting application that is more directly directed towards pedestrian support, Traunmueller & Schieck introduce a so called space recommender system [30]. There, routes can be recommended individually based on recommendations of other pedestrians concerning their walking experiences. While this system does not constitute a SUO in the narrower sense, it still can serve as a useful complement system, possibly running on public screens as recommendation input device.

Concerning public transport experience, Foell et al. propose an IoT based system for so called disadvantaged users [31]. “Disadvantaged” users are novice users, tourists, people with handicaps and older adults who have difficulties in orientating themselves properly in an unknown or uncomfortable environment. The system provides support in so-called micro-navigation, e.g. whether a person is in the correct bus or in how many minutes she needs to get off [31].

For supporting pedestrians with impaired vision Kumar et al. propose an assistance system that can run on a smart phone to detect obstacles and recognize the faces of acquaintances [32]. Although, this system shows similar problem solving structures as our system has, we rather focus to make the urban objects smart in the sense of interaction end devices.

## III. ARTIFACT DESIGN

To transform urban objects into SUOs we design a system against the following requirements: It must include

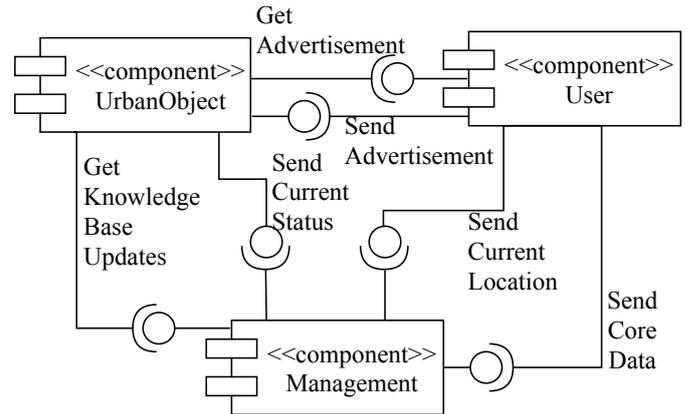


Fig. 1. Architectural Overview

- 1) the ability to book seats.
- 2) navigation information.
- 3) adaptive indications.

(1.) The ability to book seats is required to enable smart reservation of seats. (2.) Navigation information is required to guide older pedestrians via barrier-free passages. (3.) Adaptive indications are required to warn for urban hazards.

Fig. 1 shows the overall architecture of our whole system, including components for older pedestrians (*User* component) and SUOs (*UrbanObject* component). The SUO executes the *UrbanObject* component on small attached computers (e.g. “Raspberry Pi”), while mobile devices of older pedestrians (such as “Android” phones or “iPhones”) run the *User* component to allow booking requests and control Bluetooth signals. To coordinate SUOs in the overall system, we use a scalable central unit (*Management* component, see Fig. 2), based on service-oriented architecture. We use Secure Sockets Layer connections between each component and a protected database to provide basic security. Our *Management* component mainly consists of three sub-components—a seat management component (*Booking* component), urban object component (*SmartObject* component) and real time data processing component (*Live* component) for sensor data (e.g. current location of an older pedestrian or park bench temperature).

### A. Management Component

Each *User* or *UrbanObject* component manages only a subset of knowledge and cannot share knowledge with other *User* or *UrbanObject* components. Our *Management* component, therefore, provides services in distinct components to coordinate older pedestrians and SUOs and act as a global knowledge base. The following specialized components provide these services:

- *Booking*: Allocates a booking request to the best-suited seating accommodation with free seats.
- *SmartObject*: Provides knowledge and services for SUOs.
- *Live*: Integrates heterogeneous data from SUOs or devices of older pedestrians and updates the global knowledge base in real time.

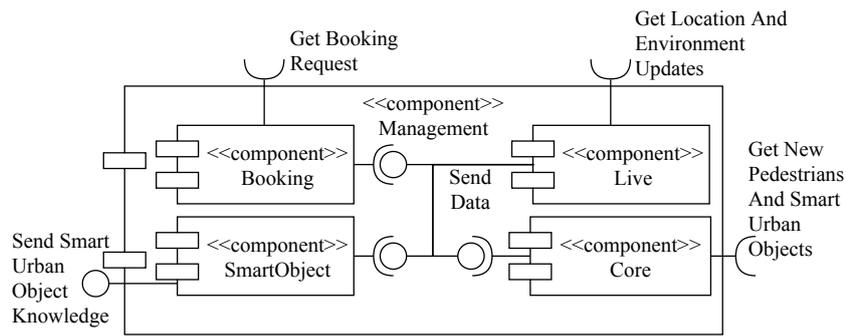


Fig. 2. The Management component

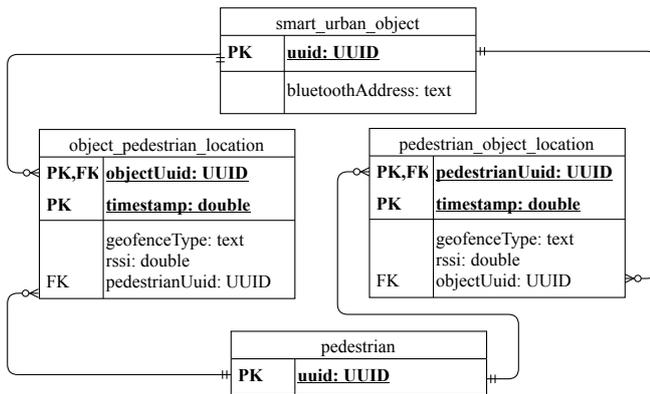


Fig. 3. Section with different primary keys

- *Core*: This component manages relatively static data such as identifiers for an older pedestrian or properties of SUOs (e.g. seat capacity).

Each of these specialized components provides REST-APIs for the *User* or *UrbanObject* component. Both components using HTTP requests to consume these REST-APIs. Hence, a network connection between components is necessary. Each component uses interchangeable data objects if they use the same data object. We're using an additional background service to manage load balance across multiple component instances and detect component failures. This background service registers all management component instances.

The interaction sequences take into account two different but simultaneous views. The first view considers the sequence from the perspective of the SUO, whereas the second considers the perspective of the older pedestrians.

### B. Interaction Sequences

The SUO sends its current status to the *UrbanObject* component and waits for the reply (see Fig. 4). Subsequently, the *SmartObject* component queries seat bookings and determines the most relevant information. This information consists either of the next seat booking on the SUO or of individual routing information for an older pedestrian. If the SUO has pending bookings, they will be transmitted, otherwise routing information for the nearest located older pedestrian. Then,

the SUO can interact with the older pedestrians, e.g. display information or adaptive indications. While older pedestrians moving through the major event area, their device uses Bluetooth Low Energy network technology to detect SUOs based on Bluetooth addresses (see Fig. 5). Subsequently, their device sends continuously RSSIs (Received Signal Strength Indicators) and Bluetooth addresses of detected SUOs to the *Live* component. Each SUO advertises Bluetooth Low Energy services with a static Bluetooth address. Based on the RSSI the *Live* component approximates the location of the older pedestrians, detects location zone changes (e.g. leaving a seat) and updates the global knowledge base. Simultaneously, the device of an older pedestrian advertise nonexistent Bluetooth Low Energy services at regular intervals (see Fig. 6). The device introduces a temporary service based on a new service identifier and a random Bluetooth address. This temporary service is not connectable for other Bluetooth devices. Meanwhile, SUOs listen to new Bluetooth Low Energy services and show adaptive light indications on urban hazards. The indication intensity depends on the highest service RSSI gathered from a listener for new service detection. When an older pedestrian approaches the SUO, visual indication on urban hazards increases. The advertisements and scans on the device of older pedestrians are independent of whether the *User* component is in the foreground or background.

### C. Live Component

SUOs and devices of older pedestrians transmit in an interval of one or two seconds, information about their environment or location. Each location consists of the recognized object and RSSI value, a timestamp and the corresponding older pedestrian. As a consequence, each location gets assigned to one SUO. The *Live* component aims to provide information to other components as quickly as possible. To avoid query join operations and time-expensive where conditions we store time-sensitive information of older pedestrians (e.g. their location or bookings) in redundant tables. Each time-sensitive information table has either the nearest SUO or an identifier for the older pedestrian as part of the primary key. Our primary key is complemented by a timestamp to read the newest insertion first, without additional sorting. Fig. 3 shows a section of our database schema providing these characteristics for

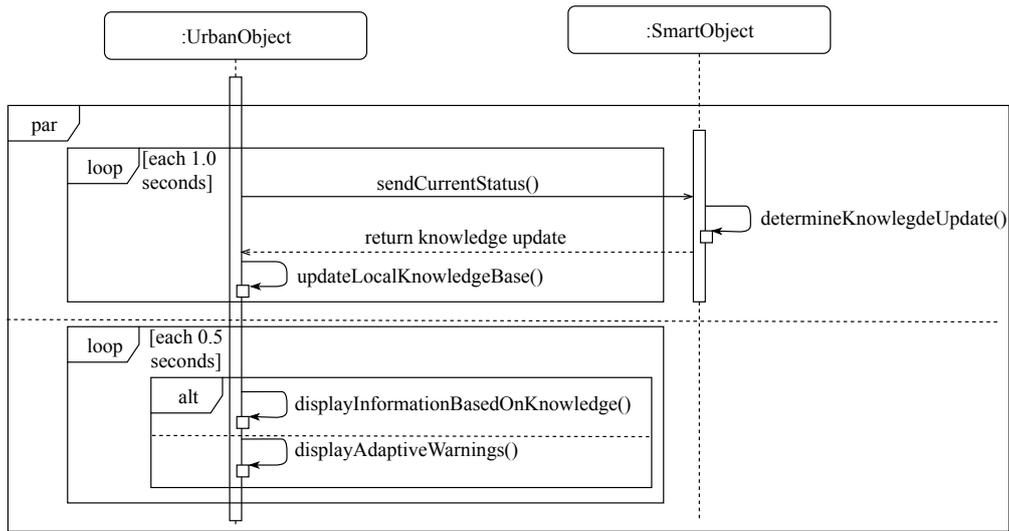


Fig. 4. Interaction between SUO and *UrbanObject* component

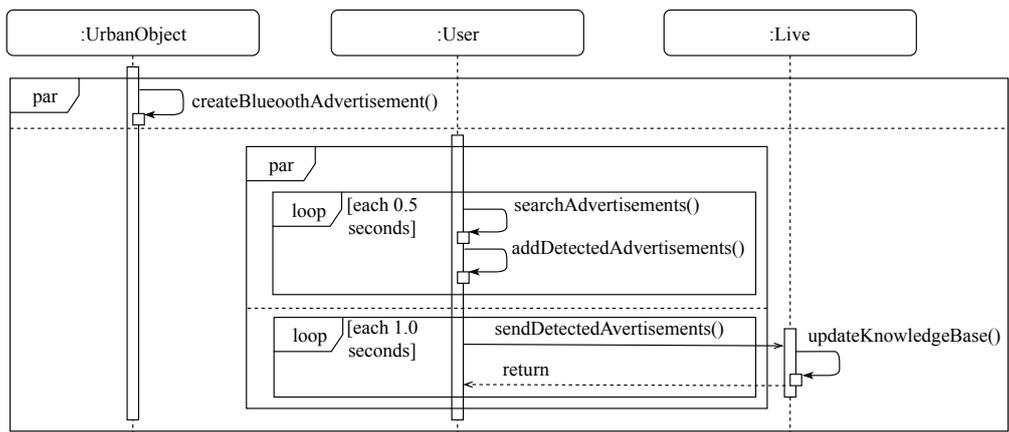


Fig. 5. Interaction between older pedestrian, SUO and real-time processing component

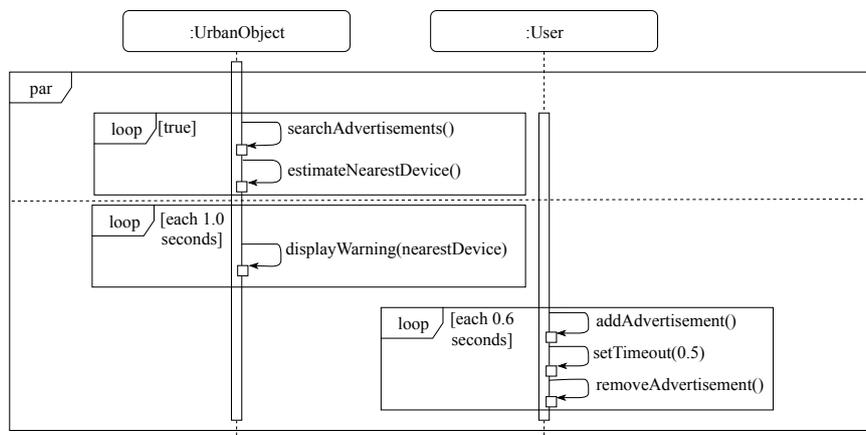


Fig. 6. Interaction between older pedestrian and SUO

locations. Due to the expected amount and frequency of data, we use a distributed database management system with two-dimensional key-value tables. Before providing information, we use gathered information to update possible dependent aspects, like the completion of bookings.

#### D. SmartObject Component

The *UrbanObject* component only receive required information for the estimated time between knowledge requests to limit resources (e.g hardware, web or computation). For this reason, the *SmartObject* component provides individual knowledge base updates for *UrbanObject* components. Its main goal is to select the most relevant information, composed of bookings on the own device or routes. We prefer information about bookings on the object to routes to other destinations for park-benches (see algorithm 1). The algorithm updates the new knowledge base with booking information of any nearby older pedestrian with a newer location update than the latest knowledge base update. If there is no older pedestrian with booking for the requesting object, our algorithm uses routing information to a booking for one older pedestrian with the latest position update as a knowledge update. To force the *UrbanObject* to show adaptive indications instead of booking or routing information, the *SmartObject* returns an empty knowledge update. In contrast to an empty knowledge update, we can force the *UrbanObject* to display independent pedestrian information by simulating bookings and locations.

---

#### Algorithm 1 Determination of $i_{new}$

---

```

 $s_r$  := Requesting SUO
 $B_r$   $\leftarrow$  Get pending and active bookings for  $s_r$ 
 $i_{new}$  := NULL;  $\triangleright$  New knowledge base
 $S$   $\leftarrow$  Set of SUOs
for  $b \in B_r$  do
   $p$   $\leftarrow$  Latest corresponding pedestrian location near to  $s_r$ 
   $p_{time}$   $\leftarrow$  Timestamp of  $p$ 
   $p_{geofence}$   $\leftarrow$  Estimate distance between  $s_r$  and  $p$ 
  if  $p_{geofence}$  is near then
    if  $p_{time}$  is newer than  $i_{new}$  then
       $i_{new}$   $\leftarrow$  Information about  $b$ 
    end if
  end if
end for
if  $i_{new} == NULL$  then
   $P$   $\leftarrow$  Pedestrian positions near  $s_r$ 
  for  $p \in P$  do
     $p_{time}$  Timestamp of  $p$ 
    if  $p_{time}$  is newer than  $i_{new}$  then
       $b$   $\leftarrow$  Next pending booking for pedestrian  $p$ 
       $b_{route}$   $\leftarrow$  Calculate barrier-free route
       $i_{new}$   $\leftarrow$  Information about  $b$ 
    end if
  end for
end if

```

---

We approximate the distance between an older pedestrian

and SUO with the obtained Bluetooth RSSI. Algorithm 2 converts the Bluetooth RSSI into an absolute number. Then, the algorithm divides the absolute number into geofences with weighting, whereby a lower geofence weight indicates a shorter distance. We use the geofence weight as distance lower bound between an older pedestrian and SUO. We use the SUO geolocation to determine pedestrians geolocation. Hence, the SUO geolocation corresponds to pedestrian geolocations if the approximated distance is close enough. Otherwise, we can't determine pedestrians geolocation.

---

#### Algorithm 2 Estimate distance $p_{geofence}$

---

```

procedure ESTIMATEDISTANCE( $p$ )
   $p_{rssi}$   $\leftarrow$  Get RSSI from a pedestrian location  $p$ 
   $p_{rssi} := \lfloor p_{rssi} \rfloor$ 
  if  $p_{rssi} < 99$  then
    return  $\lfloor \frac{p_{rssi}}{10} \rfloor$ 
  else
    return  $\infty$ 
  end if
end procedure

```

---

#### E. Booking Component

---

#### Algorithm 3 Determination of the best-suited seating accommodation $s_{opt}$

---

```

 $booking_{start}$  := Booking starting time
 $booking_{end}$  := Booking ending time
 $booking_{position}$  := Pedestrian location
 $S$   $\leftarrow$  Set of SUOs
 $s_{opt}$  := NULL;  $\triangleright$  Potential SUO with free seats.
 $s_{distance}$  := NULL;  $\triangleright$  Best-suited seat.
for  $s \in S$  do
   $B$   $\leftarrow$  Get pending and active bookings for  $s$ 
   $cap$  := Seat capacity of  $s$ 
  for  $b \in B$  do
    if  $b_{end} > booking_{start} \vee b_{start} < booking_{end}$  then
       $cap := cap - 1$ 
    end if
  end for
  if  $cap \geq 1$  then
     $distance$  := Distance between SUO and pedestrian
    if  $distance < s_{distance}$  then
       $s_{distance} \leftarrow distance$ 
       $s_{opt} \leftarrow s$ 
    end if
  end if
end for

```

---

The booking component is responsible for seat management and allocates seat preferences of older pedestrians to park benches. Its main goal is to achieve the best possible seat allocation for each older pedestrian based on their location. In order to achieve this goal, the system has to perform

two main tasks: monitoring of seats for occupancy detection and identification of the older pedestrian. It therefore uses the collected and processed data about detected SUOs by devices of older pedestrians (location of the older pedestrian), as well as information about the environment of seating accommodations. Algorithm 3 shows the allocation algorithm using environment and information of older pedestrians. The algorithm validates for each pending and active bookings of all seating accommodations if any seating accommodations have seats for the booking request. We calculate the number of free seats for one seating accommodations through subtraction of reserved bookings in the corresponding time slot from the individual total number of seats. This time slot starts five minutes and ends 30 minutes after the booking request. Thereby, this results in a 30-minute seat booking. We use the estimated distance between older pedestrians and seating accommodation as decision base and prefer a short distance. Therefore, our algorithm provides the nearest free seat accommodation. Subsequently, we transmit the allocated seat to our *Live* and *User* component to enable routing and provide visual feedback. If the older pedestrian arrives his booked park bench and her *User* component recognizes the seating accommodation at least two times, we mark the booking as active. We mark the seat as free if the older pedestrian leaves the seating accommodation or his booking end occurs. The older pedestrian leaves the seating accommodation if her latest obtained location is in an outer geofence zone for this seating accommodation.

#### F. Core Component

Before interacting with older pedestrians, the *User* and *UrbanObject* components need to register once and become known to other component instances. Hence, the *Core* component creates a unique random identifier for each new *User* or *UrbanObject* and informs the *Live* component. Another responsibility for this component is the attribute management for different kinds of SUOs. The *UrbanObject* component provides additional attributes to identify it in case of downtime, whereas the *User* component identifies itself (e.g. name or geographic coordinates). If any *User* component can not identify itself, we consider this *User* as new *User* component. Further attributes for components depend on the corresponding object type. This includes seat capacity for SUOs with type seating accommodation, whereas the type for urban hazards include different indication types.

#### G. UrbanObject Component

The *UrbanObject* component on the SUO control detection of an older pedestrian and provide visual feedback for the older pedestrian. Therefore, we connect the component with a color display and Bluetooth Low Energy Module. We overwrite obtained knowledge from *SmartObject* component and recognized older pedestrian if newer knowledge is available. If knowledge is available, the *UrbanObject* component can show this knowledge. Due to random Bluetooth address from mobile devices of older pedestrians, we aggregate recognized older

pedestrians. Then, we submit this aggregation as status to *Live* component. If the component shows an adaptive indication, we transform the highest Bluetooth RSSI within two seconds into a percentage value and use this value as indication intensity. In case of a difference greater ten between the last and current percentage value, we use the average as indication intensity for a smooth transition.

#### H. User component

Apart from a user interface to request a seat, the *User* component is responsible for the localization of an older pedestrian. Localization consists of scanning for *SmartObject* components and advertising of own services. This component compares detected Bluetooth devices against Bluetooth addresses in its knowledge base and filters *UrbanObject* components. Due to possible operating system restrictions from mobile devices of older pedestrians and changed settings, the *User* component has to monitor the outcome of localization operations. If the *User* component detects any deviation, it pauses the concerning operation, until the user solves it. We store the unique random identifier for an older pedestrian on their mobile device to keep the same identifier and allow component shutdowns and restarts. Thus, after a restart, it starts automatically scanning and advertising. If an older pedestrian rejects the localization, he enables an incognito mode, where his device advertises services for adaptive indication but do not scan for *UrbanObject* components. Consequently, no localization information will be transmitted. The mobile device has to support Bluetooth Low Energy to use the overall system.

## IV. FUNCTIONAL TESTING

To show the feasibility of our artifact, we conducted two scenarios, addressing different safety aspects. The first scenario, shown in figure 7, addresses the booking of resting possibilities to counteract against exhaustion. In contrast to the first scenario, scenario two (see fig. 8) focuses on the prevention of risks from urban hazards by adaptive indications. Scenario “seating accommodation” covers the need for seats based on exhaustion or physical restrictions of an older pedestrian if elderly recognize their need. Then, they announce their need for a seat, in this case, to our artifact. Our artifact will search for the best-suited seat, books this seat and informs sub-components. Subsequently, the older pedestrian receives her booking and goes to her seat. On the way, she sees routing information from *UrbanObject* components that allows her to find the barrier-free way to his seat. When arriving, the seat shows that the older pedestrian reached his destination.

The second scenario “adaptive indication” shows indications in different intensity based on the estimated distance between urban hazards and older pedestrians. On the other hand, if the older pedestrian walk away, the indication intensity decreases.

We suppose our artifact to be feasible if we provide technical functionality that is required to enable seat bookings for resting possibilities and adaptive indications on urban hazards.

Therefore, our artifact has to approximate the distance between the older pedestrian and urban hazards, estimate the

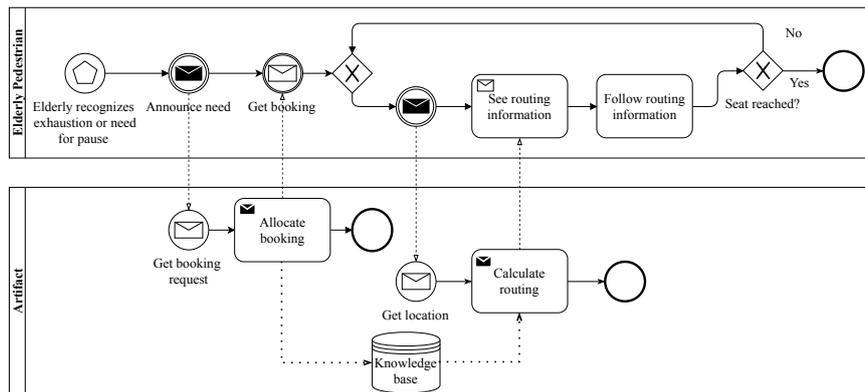


Fig. 7. Scenario “seating accommodation”

location of the older pedestrian and distribute knowledge. We validate the location detection and knowledge distribution in scenario “seating accommodation” and use scenario “adaptive indication” to prove the approach detection.

Our scenario-based testing takes primary place on the property of our institute. However, we have been able to reproduce the settings at the “Turmfest Rheydt 2019”, a major event in a larger German city. The settings include Raspberry Pi’s with LED matrices, Android phones, a laptop and a router. We have positioned Raspberry Pi’s at least five meters apart. Each Raspberry Pi runs the *UrbanObject* component with a preregistered SUO, while the laptop executes the *Management* component with sub-components. Android phones run the *User* component that includes a preregistered older pedestrian. Each device connects to a private wireless network, made available by a router, to consume REST-APIs. We implement the *UrbanObject* component in Python, whereas the *User* component is in Typescript and the *Management* component in Java. Our *Management* component uses Apache Cassandra as database system. In the study are other Bluetooth Low Energy devices, such that *UrbanObject* components started to run two minutes before starting each scenario.

In scenario “seating accommodation”, the Raspberry Pi’s represented seating accommodations. We requested seats via the *User* component, while we stayed outside the detection range of all Raspberry Pi’s. Then, we got a seat allocation and started to walk to a component, which was not our allocated seating accommodation. The *UrbanObject* component on this Raspberry Pi displayed an arrow towards our allocated virtual seating accommodation within five seconds. Meanwhile, we went further to our allocated seating accommodation, that displayed a symbol that represents a free seat when arrived in a zone of five meters. If entered a zone of a half meter and allocated seating accommodation, the component displays another different symbol within ten seconds. The scenario shows that our artifact can estimate locations of an older pedestrian and distribute knowledge across multiple SUOs. However, there can be a time gap of five to ten seconds between arriving and detection of arriving.

In contrast to the scenario “seating accommodation”, in

scenario “adaptive indication” our Raspberry Pi’s simulated urban hazards. Again, we started outside the detection range of all Raspberry Pi’s, so no indication was displayed. We went in walking pace straight towards a Raspberry Pi. The indication appeared during the movement and the warning intensity increased with decreasing distance. Finally, the indication disappeared. However, the indication intensity jumped at the same distance. We find the Bluetooth RSSI of the Android phone changed in a range of plus-minus six at the same distance. In fact, the scenario “adaptive indication” demonstrates our artifact provides approach detection. Nevertheless but we can not provide the exact mapping from Bluetooth RSSI to distance. Based on findings of scenario “seating accommodation” and scenario “adaptive indication” we suppose the feasibility of our artifact.

## V. DISCUSSION

We contribute a system to enable adaptive indications of urban hazards, barrier-free passages and smart reservation of seats. Thus, we transform urban objects into connected SUOs. We described two scenarios where SUOs support elderly to participate in major events. Our evaluation does not apply to adapted scenarios that include shielding elements or smaller distances between SUOs. In contrast, our evaluation applies to scenarios in urban areas or with an increased number of SUOs. We did not consider or compare the signal strengths of different small computers (inclusive accessories like displays) and mobile phones, including their operating system. Findings concerning the processing or handling of volatile Bluetooth RSSI shows white spots for further research. The artifact extension of localization with GPS-geographic coordinates may be future work. Due to our centralized approach, scalability limitations may occur. As a consequence, we can distribute our database about multiple machines or instantiate multiple component instances to counter.

Since our artifact does not require any special characteristics of the elderly and is not tied to any position or event type, our artifact can be used by all pedestrians and in all urban areas. The smart seat reservation including routing via barrier-free passages corresponds to routing from one location to another.

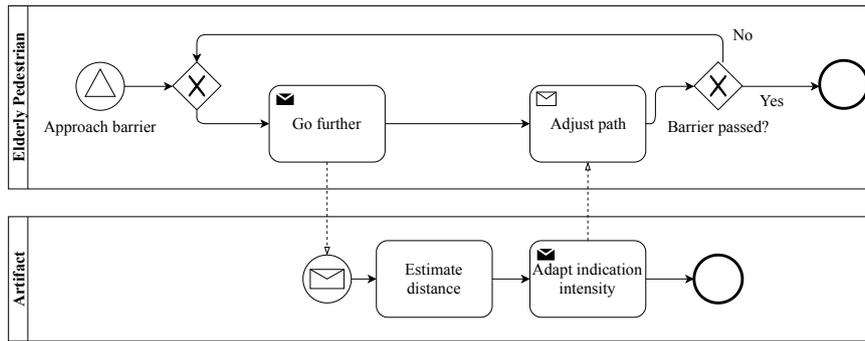


Fig. 8. Scenario “adaptive indication”

In particular, people with a handicap can benefit from our artifact, too. Included functions in our artifact are transferable to other use cases. For example, the smart reservation of rest possibilities can constitute a basis for individual routing to and reservation of toilets. Another application may be the smart reservation of parking slots (including routing) or the barrier-free finding of dining possibilities and dine pre-order for the elderly for major events. In addition, our artifact can support pedestrians in their everyday lives. The consideration of environment data (e.g. the number of people in a particular area) for routing calculation may be future work too.

Regarding our understanding of sensors and actuators explicated in section II, our SUOs can be classified as I-S-A-N objects according to the typology of López [22]. The SUOs have a unified identifier and some data storage (I), are able to sense the environment through Bluetooth signals (S), can affect the environment with visually outputs from the LED displays (A) and are connected in a network (N). For our purposes, this network can be a local area one. However, connecting the SUOs to the internet would be conceptually equal. Decision-making ability (D), that goes beyond simple stimulus-response, is delegated to the central management component, cf. figure 2. This management component implements the service layer in the sense of Xu et al. [20], as described in section II, too.

Revisiting our research question, table I summarizes how our SUOs can contribute to safety. The safety-engineering approach guided our pre-design phases with the conceptual separation of hazards and worst-case environmental condition. By “accepting” that we cannot alter some circumstances concerning the pedestrian state or the environmental conditions, we focused on means for avoiding that worst-case environmental condition *coincide* with an older pedestrian being in a defined hazardous state. Therefore it is of less importance whether the pedestrians are defined to be the system or the SUOs than to actually define respective states and conditions. These need to be defined in a way such that an engineer gets a “point of attack”. When a hazard is given, the engineer can try to find a way, how the environment can be adapted with feasible means. When an environmental condition is given, the engineer can try to find a way, how the system resp. pedestrian

state can be adapted, if possible. As adaption of the pedestrian state is seldom possible, a feasible way is then to find means that the pedestrian avoids sub-environments with worst-case conditions. Quite obviously, in our case we could also have referred to worst case environmental conditions as hazards in the environmental subsystem. We decided to define the pedestrians to be the system because our explicit approach is to make the *environment* adaptive to the pedestrians requirements.

## VI. CONCLUSIONS

We proposed Smart Urban Objects (SUOs) for (1.) adaptive indication of accessible passages, (2.) adaptive indication of stumbling blocks and for (3.) reservation of seat on public outdoor major events. Example scenarios are respectively:

- 1) An older pedestrian approaches a step but cannot lift her legs high enough to safely pass it. The step as an SUO detects the older pedestrian and points in the direction of an accessible passage with a lower step or no step at all.
- 2) An older pedestrian approaches a “hidden” stumbling block, like a small polder. The polder as an SUO detects the older pedestrian and warns her visually. Thereby we expect that an adaptive warning will be more salient and hence contributes more perceived safety of older pedestrians than a static one, which is always visible.
- 3) An older pedestrian needs a seat rest. She can reserve a seat from anywhere on the area. When she passes another SUO, the display of the SUO shows a pointer in the direction of the reserved seat. When the older pedestrian approaches the reserved seat, the seat’s visual reservation signal indicates that the pedestrian can sit down there.

With our SUOs we seek to enhance safety for older pedestrians who often have physical impairments and reduced resilience. The SUOs rely on detection of the older pedestrians via Bluetooth technology and provide output with LED-displays. The SUOs are implemented with Raspberry Pi.

We conducted scenario based functional testing to validate whether our SUOs are feasible with standard Bluetooth technology. The SUOs have been installed at a major event in a larger German city and have been positively tested with respect to its functionality in the target environment.

TABLE I  
CONTRIBUTION TO SAFETY WITH SUOS

Hazardous pedestrian state	Worst-case environmental condition	Contribution of SUO in avoiding the coincidence of hazardous pedestrian state and worst-case environmental condition
Inability to overcome structural barriers	Structural barrier on passage	SUO adaptively point in the direction of an accessible passage.
Pedestrian doesn't recognize obstacles	Stumbling block in pedestrian's proximity	SUO adaptively indicates the existence of a stumbling block.
Pedestrian needs a seat rest	No awareness on seat availability	Reservation of a seat from anywhere is possible. Adaptive signs guide the pedestrian to the reserved seat.

#### ACKNOWLEDGEMENT

This work has been supported by the Federal Ministry of Education and Research, Germany, under grant 16SV7438K.

#### REFERENCES

- [1] OECD, Ed., *Ageing and Transport*. Paris: OECD, 2001.
- [2] J. D. Moreland, J. A. Richardson, C. H. Goldsmith, and C. M. Clase, "Muscle Weakness and Falls in Older Adults: A Systematic Review and Meta-Analysis," *Journal of the American Geriatrics Society*, vol. 52, no. 7, pp. 1121–1129, 2004. doi: 10.1111/j.1532-5415.2004.52310.x
- [3] N. D. Carter, P. Kannus, and K. M. Khan, "Exercise in the Prevention of Falls in Older People," *Sports Medicine*, vol. 31, no. 6, pp. 427–438, 2001. doi: 10.2165/00007256-200131060-00003
- [4] A. M. Bastos, C. G. Faria, E. Moreira, D. Morais, J. M. Melo-de Carvalho, and M. C. Paul, "The importance of neighborhood ecological assets in community dwelling old people aging outcomes: A study in Northern Portugal," *Frontiers in Aging Neuroscience*, vol. 7, p. Article 156, 2015. doi: 10.3389/fnagi.2015.00156
- [5] J. D. Fortuijn, M. van der Meer, V. Burholt, D. Ferring, S. Quattrini, I. R. Hallberg, G. Weber, and G. C. Wenger, "The activity patterns of older adults: a cross-sectional study in six European countries," *Population, Space and Place*, vol. 12, no. 5, pp. 353–369, 2006. doi: 10.1002/psp.422
- [6] Z. Gabriel and A. Bowling, "Quality of life from the perspectives of older people," *Ageing and Society*, vol. 24, no. 5, pp. 675–691, 2004. doi: 10.1017/S0144686X03001582
- [7] J. Leukel, B. Schehl, S. Wallrafen, and M. Hubl, "Impact of IT Use by Older Adults on Their Outdoor Activities," in *Proc. of the 38th Int. Conf. on Information Systems (ICIS 2017)*, 2017.
- [8] T. Hausteijn, J. Mischke, F. Schönfeld, and I. Willand, Eds., *Older people in Germany and the EU*. Wiesbaden: Federal Statistical Office, 2016.
- [9] M. Rantakokko, S. Iwarsson, M. Kauppinen, R. Leinonen, E. Heikkinen, and T. Rantanen, "Quality of Life and Barriers in the Urban Outdoor Environment in Old Age," *Journal of the American Geriatrics Society*, vol. 58, no. 11, pp. 2154–2159, 2010. doi: 10.1111/j.1532-5415.2010.03143.x
- [10] T. Sugiyama and C. W. Thompson, "Environmental Support for Outdoor Activities and Older People's Quality of Life," *Journal of Housing For the Elderly*, vol. 19, no. 3-4, pp. 167–185, 2006. doi: 10.1300/J081v19n03\_09
- [11] F. Bellotti, E. Ferretti, and A. De Gloria, "Discovering the European Heritage Through the ChiKho Educational Web Game," in *Proc. of the 1st Int. Conf. on Intelligent Technologies for Interactive Entertainment (INTETAIN 2005)*. Berlin, Heidelberg: Springer, 2005, pp. 13–22.
- [12] E. S. Poulsen, A. Morrison, H. J. Andersen, and O. B. Jensen, "Responsive lighting," in *Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services - MobileHCI '13*. ACM Press, 2013. doi: 10.1145/2493190.2493218 p. 217.
- [13] P. Cremonesi, A. D. Rienzo, and F. Garzotto, "Personalized interactive public screens," in *Proc. of the 4th Workshop on Interacting with Smart Objects (at ACM IUI 2015)*, 2015, pp. 10–15.
- [14] J. Yang, J. Portilla, and T. Riesgo, "Smart parking service based on Wireless Sensor Networks," in *IECON 2012 - 38th Annual Conference on IEEE Industrial Electronics Society*. IEEE, 2012. doi: 10.1109/IECON.2012.6389096 pp. 6029–6034.
- [15] R. J. Shepard, "Age and Physical Work Capacity," *Experimental Aging Research*, vol. 25, no. 4, pp. 331–343, 1999. doi: 10.1080/036107399243788
- [16] U.S. Department of Transportation, *Improving Transportation for a Maturing Society*. Washington, DC: Office of the Assistant Secretary for Transportation Policy, 1997.
- [17] N. G. Leveson, *Engineering a Safer World*. Massachusetts: The MIT Press, 2012.
- [18] A. Rayes and S. Salam, *Internet of Things From Hype to Reality*. Cham: Springer International Publishing, 2019.
- [19] L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey," *Computer Networks*, vol. 54, no. 15, pp. 2787–2805, 2010. doi: 10.1016/j.comnet.2010.05.010
- [20] L. D. Xu, W. He, and S. Li, "Internet of Things in Industries: A Survey," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 4, pp. 2233–2243, 2014. doi: 10.1109/TII.2014.2300753
- [21] G. Kortuem, F. Kawasar, D. Fitton, and V. Sundramoorthy, "Smart objects as building blocks for the internet of things," *IEEE Internet Computing*, vol. 14, no. 1, pp. 44–51, 2010. doi: 10.1109/MIC.2009.143
- [22] T. S. López, D. C. Ranasinghe, B. Patkai, and D. McFarlane, "Taxonomy, technology and applications of smart objects," *Information Systems Frontiers*, vol. 13, no. 2, pp. 281–300, 2011. doi: 10.1007/s10796-009-9218-4
- [23] J.-P. Vasseur and A. Dunkels, "Smart Cities and Urban Networks," in *Interconnecting Smart Objects with IP*. Burlington: Elsevier, 2010, ch. 22, pp. 335–351.
- [24] B. Hammi, R. Khatoun, S. Zeadally, A. Fayad, and L. Khoukhi, "IoT technologies for smart cities," *IET Networks*, vol. 7, no. 1, pp. 1–13, 2018. doi: 10.1049/iet-net.2017.0163
- [25] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of Things for Smart Cities," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 22–32, 2014. doi: 10.1109/JIOT.2014.2306328
- [26] A. Huldgren, C. Katsimerou, A. Kuijsters, J. A. Redi, and I. E. J. Heynderickx, "Design Considerations for Adaptive Lighting to Improve Senior's Mood," in *Proceedings of the 13th International Conference on Smart Homes and Health Telematics (ICOST 2015)*. Springer (LNCS), 2015. doi: 10.1007/978-3-319-19312-0\_2 pp. 15–26.
- [27] M. Cunha and H. Fuks, "AmbLEDs: Implicit I/O for AAL Systems," in *Proceedings of the 4th Workshop on Interacting with Smart Objects (at ACM IUI 2015)*, 2015, pp. 6–9.
- [28] J. Müller, F. Alt, D. Michelis, and A. Schmidt, "Requirements and design space for interactive public displays," in *Proceedings of the international conference on Multimedia - MM '10*. New York, New York, USA: ACM Press, 2010. doi: 10.1145/1873951.1874203 p. 1285.
- [29] D. Vogel and R. Balakrishnan, "Interactive public ambient displays," in *Proceedings of the 17th annual ACM symposium on User interface software and technology - UIST '04*. ACM Press, 2004. doi: 10.1145/1029632.1029656 p. 137.
- [30] M. Traunmueller and A. Fatah gen. Schieck, "Introducing the space recommender system," in *Proceedings of the 6th International Conference on Communities and Technologies - C&T '13*. ACM Press, 2013. doi: 10.1145/2482991.2482995 pp. 149–156.
- [31] S. Foell, G. Kortuem, R. Rawassizadeh, M. Handte, U. Iqbal, and P. Marrón, "Micro-Navigation for Urban Bus Passengers: Using the Internet of Things to Improve the Public Transport Experience," in *Proceedings of the The First International Conference on IoT in Urban Space*. ICST, 2014. doi: 10.4108/icst.urb-iot.2014.257373 pp. 1–6.
- [32] P. M. Kumar, U. Gandhi, R. Varatharajan, G. Manogaran, J. R., and T. Vadivel, "Intelligent face recognition and navigation system using neural learning for smart security in Internet of Things," *Cluster Computing*, 2017. doi: 10.1007/s10586-017-1323-4

# Advances in Information Systems and Technology

**I**ST is a FedCSIS conference track aiming at integrating and creating synergy between FedCSIS technical sessions that thematically subscribe to the disciplines of information technology and information systems. The track emphasizes the issues relevant to information technology and necessary for practical, everyday needs of business, other organizations and society at large. This track takes a sociotechnical view on information systems and relates also to ethical, social and political issues raised by information systems. Technical sessions that constitute IST are:

- AITM'19—16<sup>th</sup> Conference on Advanced Information Technologies for Management
- DSH'19—1<sup>st</sup> Special Session on Data Science in Health
- InC2Eco'19—1<sup>st</sup> Workshop on Data Analysis and Computation for Digital Ecosystems
- ISM'19—14<sup>th</sup> Conference on Information Systems Management
- KAM'19—25<sup>th</sup> Conference on Knowledge Acquisition and Management



# 17<sup>th</sup> Conference on Advanced Information Technologies for Management

**W**E are pleased to invite you to participate in the 17<sup>th</sup> edition of Conference on “Advanced Information Technologies for Management AITM’19”. 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

- 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

- **Andres, Frederic**, National Institute of Informatics, Tokyo, Japan
- **Dudycz, Helena**, Wrocław University of Economics, Poland
- **Dyczkowski, Mirosław**, Wrocław University of Economics, Poland
- **Hunka, Frantisek**, University of Ostrava, Czech Republic
- **Korczak, Jerzy**, International University of Logistics and Transport, Wrocław, Poland

## PROGRAM COMMITTEE

- **Abramowicz, Witold**, Poznan University of Economics, Poland
- **Ahlemann, Frederik**, University of Duisburg-Essen, Germany
- **Atemezing, Ghislain**, Mondeca, Paris, France
- **Cortesi, Agostino**, Università Ca’ Foscari, Venezia, Italy
- **Czarnacka-Chrobot, Beata**, Warsaw School of Economics, Poland

- **De, Suparna**, University of Surrey, Guildford, United Kingdom
- **Dufourd, Jean-François**, University of Strasbourg, France
- **Franczyk, Bogdan**, University of Leipzig, Germany
- **Januszewski, Arkadiusz**, University of Science and Technology, Bydgoszcz, Poland
- **Kannan, Rajkumar**, Bishop Heber College (Autonomous), Tiruchirappalli, India
- **Kersten, Grzegorz**, Concordia University, Montreal, Canada
- **Kowalczyk, Ryszard**, Swinburne University of Technology, Melbourne, Australia
- **Kozak, Karol**, TUD, Germany
- **Krótkiewicz, Marek**, Wrocław University of Science and Technology, Poland
- **Leyh, Christian**, University of Technology, Dresden, Germany
- **Ligeza, Antoni**, AGH University of Science and Technology, Poland
- **Ludwig, André**, Kühne Logistics University, Germany
- **Magoni, Damien**, University of Bordeaux – LaBRI, France
- **Michalak, Krzysztof**, Wrocław University of Economics, Poland
- **Owoc, Mieczyslaw**, Wrocław University of Economics, Poland
- **Pankowska, Malgorzata**, University of Economics in Katowice, Poland
- **Pinto dos Santos, Jose Miguel**, AESE Business School Lisboa, Portugal
- **Proietti, Maurizio**, IASI-CNR (the Institute for Systems Analysis and Computer Science), Italy
- **Rot, Artur**, Wrocław University of Economics, Poland
- **Stanek, Stanislaw**, General Tadeusz Kosciuszko Military Academy of Land Forces in Wrocław, Poland
- **Surma, Jerzy**, Warsaw School of Economics, Poland and University of Massachusetts Lowell, United States
- **Tazi, El Bachir**, Moulay Ismail University, Meknes, Morocco
- **Teufel, Stephanie**, University of Fribourg, Switzerland
- **Tsang, Edward**, University of Essex, United Kingdom
- **Wątróbski, Jarosław**, University of Szczecin, Poland
- **Weichbroth, Paweł**, WSB University of Gdansk, Poland
- **Wendler, Tilo**, Hochschule für Technik und Wirtschaft Berlin

- **Wolski, Waldemar**, University of Szczecin, Poland
- **Zanni-Merk, Cecilia**, INSA de Rouen, France

- **Ziemia, Ewa**, University of Economics in Katowice, Poland

# Factors Influencing the Intended Adoption of Digital Transformation: A South African Case Study

Rion van Dyk

Department of Information Systems,  
University of Cape Town  
P Bag, Rondebosch, South Africa  
rionvandyk@gmail.com

Jean-Paul Van Belle

CITANDA, Department of Information Systems,  
University of Cape Town  
P Bag, Rondebosch, South Africa  
Jean-Paul.VanBelle@uct.ac.za

**Abstract**—Organisations worldwide are facing a market pressures which are forcing them to undertake digital transformation projects or initiatives. This research study set out to explore this in more depth, asking questions about the intention of South African (SA) retail organisations to adopt digital transformation and probing into the understanding and perception of digital transformation itself, as well as future intended use cases for available digital technologies. A case study approach was deemed to be an appropriate strategy given the paucity of existing academic research on the topic. Participants in the study had a good understanding of digital transformation and digital technologies, but perceptions were mixed. The most prominent digital transformation initiatives in the SA retail industry were the adoption of cloud technologies and data analytics. The factors identified in this study, using the Technology, Organisational, and Environmental (TOE) framework, can assist retailers in their decision-making process concerning digital transformation adoption.

**Key Words:** Digital Transformation, Digital Technologies, Strategy, South African Retail Organisations, Technology Adoption, Perceptions, Understanding, TOE, Technology Drivers.

## I. INTRODUCTION

Many industries have been facing a market shift over the past few years driven by a better response to customer demand which forced enterprises to undertake digital transformation projects or be left behind the competition [1]. Digital transformation refers to a type of strategy that changes an enterprise business model which ultimately provides the customer with variants of tangible product by taking advantage of new or existing digital technologies [2]. The competitive landscape is changing in many industries due to business digitalization. Enterprises face threats of digital disruption from new market entrants while digitally savvy customers are demanding more from the enterprise [3]. Digital transformation affects every enterprise and sector as the market-changing potential of digital technologies is often wider than sales channels, supply chains, products and business processes [4].

One of the biggest challenges enterprises currently face is integrating and exploiting new digital technologies [4].

Digital technologies are tools that enterprises must make use of to get closer to their customers, transform their business processes and empower their employees [5]. Current new digital technologies include cloud computing, mobile, analytics, social media, robotics and Internet of Things (IoT) technologies [6]. These digital technologies can present the enterprise with game-changing opportunities if they are combined with accessibility of enterprise data to enrich their products, services and customer relationships [7].

There is a lack of information around digital transformation, its perceptions and use cases in the SA retail industry to aid its adoption. The main objective of this study was to understand and examine the current perceptions and status of digital transformation within a SA retail organisation. Furthermore, the study aimed to identify factors influencing the intended adoption of digital transformation within the SA retail organisation. This will provide the information and knowledge needed for the retail industry to make informed decisions about the potential future use of digital technologies and how to overcome adoption barriers.

The following propositions are posed to relate the research findings to existing theories and models identified in the literature review.

- PR1: There is a lack of understanding of digital transformation in the SA retail industry and perceptions of digital transformation are mixed.
- PR2: SA retail industry organisations have identified specific core technologies driving digital transformation.
- PR3: The intended adoption of digital transformation by SA retail organisations is affected by specific TOE (Technology, Organisational, & Environmental) factors.

It is important to study the factors influencing the intended adoption of digital transformation so that enterprises can understand the challenges and address them. Addressing these challenges will be beneficial to the enterprise as it will assist it to create a clear and coherent digital strategy, lead to retaining and attracting top talent, and create a company culture where employees could be innovative and creative. Ultimately, a digital transformed enterprise will be able to easily adapt taking advantage of new opportunities and have a competitive advantage over their competition.

## II. LITERATURE REVIEW

### A. Digital Transformation

Digital transformation refers to an enterprise business model that applies new or existing digital technologies and products or services into digital variants to offer a tangible product to their customers [2].

Digital transformation is not only about technology, but it also requires a new way of thinking and strategy by enterprise executives. *“Digital transformation is the profound transformation of business and organisational activities, processes, competencies and models to fully leverage the changes and opportunities of a mix of digital technologies and their accelerating impact across society in a strategic and prioritized way, with present and future shifts in mind”* [1]. Enterprise digital transformation strategies should include the application of digital technologies to enterprise processes, products and assets to enhance customer value, uncover new monetization opportunities, improve efficiencies and manage risk across the enterprise [8] [9].

### B. Digital Technologies

New digital technologies (social, mobile, analytics, cloud computing and Internet of Things [IoT] technologies) could present the enterprise with game-changing opportunities and existential threats. Leaders in digital transformation apply new digital technologies and related technologies in conjunction with the accessibility of enterprise data to enrich their products, services and customer relationships [7].

#### Social Media

The phenomenal and exponential growth of social media and mobile resulted in many organisations realizing that an online presence is required to reach out and connect with their digital savvy customers [10]. Capturing data from tools such as Facebook, LinkedIn and blogs are essential to integrate the information into the sales process [11]. Digital savvy customers follow brands on social media and expect to be able to view store inventory online to enable them to do “showroom” shopping before going into a physical store [10].

#### Mobility

Digital technologies have enabled enterprises to make use of mobility and ubiquitous connectivity features providing the enterprise with immediate interaction and access to a wide range of data and computing power and enabling enterprises to analyse their data and make decisions in real time [12]. Mobility resulted that tech-savvy customers across all facets of society completely changed their behaviours, expectations and the way they interact with enterprises [4].

Mobile penetration throughout Africa often thought of as the least digitally populated continent has reached 70% of its one billion inhabitants. Over 40% of the world’s population has an internet connection. Mobile technology advances allow for the capture of geographical and contextual data that was previously not possible [13]. Digitalization experienced a significant boost with the introduction of smart mobile devices and the applications that run on them [14].

Furthermore, the declining cost of mobile technologies has broadened their potential for worldwide use [13].

#### Analytics

Digital technologies provide enterprises with interpretation capabilities, enabling in-depth analysis and exploration of different kinds of data sets. Digital analytical tools coupled with computer-enabled techniques can yield insight to enterprise executives from massive multidimensional datasets enabling them to make use of analytics to make strategic enterprise decisions [13]. About 90% of the data available today has been produced in the last two years. This data explosion has been driven by new data sources such as digital transactions, mobile devices, embedded sensors and the growing use of social media by the global population. Enterprises could benefit from learning how to capture, absorb, store and analyse their data and turn their data into a valuable asset [14].

Data analytics should be incorporated into new digital products for personalisation reasons, but also to inform other enterprise departments like product development, sales and marketing [15]. Some enterprises are known for their analytical-based approach to ensure they personalise their service and marketing to the need of each of their individual customers by constantly innovating, improving their processes, launching new service based data-driven applications and capabilities [16].

#### Cloud Computing

Cloud computing has been defined as *“a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction”* [17]. Cloud computing technologies enable enterprises to outsource some elements of the IT value chain with benefits for enterprises such as reduced costs, scalability, flexibility, capacity utilisation, higher efficiencies and mobility [18].

#### Internet of Things (IoT)

The Internet of things (IoT) refers to a type of network that enables any device to connect to the internet based on stipulated protocols through information sensing equipment to conduct communication and information exchange. The IoT concept has become more practical in recent years due to the exponential growth of the use of smart mobile devices, the growth of data analytics and cloud computing. IoT enables things to be connected at anytime, anywhere, with anything and with anyone ideally using any network, path or service [19]. IoT will force enterprises to digitally transform and will bring fundamental changes to individuals’ and society’s expectation and perspectives on how technologies and applications work in the world [20].

C. Adoption Barriers

Digital Strategy

One of the biggest barriers to digital transformation and digital maturity is the lack of a clear and coherent digital strategy to drive transformation within the enterprise. Mature digital companies realise that digital technologies should be used to achieve strategic enterprise goals [9]. Digital strategy can be defined as “a business strategy, inspired by the capabilities of powerful, readily accessible digital technologies, intent on delivering unique, integrated business capabilities in ways that are responsive to constantly changing market conditions” [7]. Executives should use the enterprise digital strategy to create competitive advantage, value and customer satisfaction by combining existing technology with capabilities of other digital technologies [7].

Talent

Employing, retaining and developing talent within the enterprise is another challenge faced by digital maturing enterprises [9]. Talent was highlighted as one of the top three key influencing factors of successfully implementing digital transformation projects [21]. Having the right talent allows the enterprise to adapt to change and create new opportunities [9]. Digital leaders should embrace the fact that talented individuals are on the lookout for the best digital opportunities and want to work for digitally enabled enterprises [9].

Company Culture

Digital mature enterprises need to foster a culture where employees are encouraged to take risks, innovate, be creative and create a collaborative work environment [9]. Trying out a lot of things and learning quickly from errors can only be done if a culture of trial-and-error exists within the enterprise [15]. Changing the company culture is a real challenge during digital transformation. Traditional corporate cultures present executives with resistance to change and barriers that need to be overcome to ensure successful digital transformation [22].

Leadership

A lack of strong top-down leadership that steers digital transformation by setting direction, building momentum and ensuring the enterprise follows through is another barrier to becoming digitally transformed [5]. Enterprise executives are advised to focus on employees, culture, talent, skillset and leadership [12]. Enterprise executives must drive and coordinate digital transformation across the entire enterprise as it touches every functional area. The entire workforce must “digitalize” by fostering an enterprise culture of collaboration between departments whereby employees are enabled to exchange opinions and ideas across departments [15].

IT Function Transformation

Digital transformation has prompted enterprise IT departments to rethink the architecture platforms within the organisation to ensure new business demands originating from customers and suppliers who require more digital engagement are met [21]. Many IT departments are not set up to be flexible and agile to ensure quick and fast modifications to

applications on short notice when required by business departments. Flexibility, agility and the ability to service business requirements on short notice are all trademarks of a digitally transformed enterprise. The role of IT service provider needs to change to a role of consultant, enabler and innovator by applying “new IT concepts” like cross-functional digital teams, enterprise architecture (EA), co-location and IT innovation management [23].

Omni-channel (OC)

The customer retail experience has been profoundly transformed in the past 10 years by digital transformation projects which integrate digital technologies into the customer shopping experience [24]. Physical stores remain an important part of the customer shopping experience even though mobile device sales and online commerce are accelerating. But the distinction between traditional brick and mortar stores and online sales channels will disappear as the retail industry moves into a new phase, known as Omni-channel (OC) retailing [25]. OC retailing can be defined as the process whereby customers are influenced and move through multiple digital channels in their search and buying process [25].

D. Technology, Organisation, Environment framework (TOE)

The TOE framework was proposed by [26]. It “represents how different elements of an organisation (technology, organisation and environment) affect technological innovations and that the framework is suitable for research as it has the flexibility for variance of the factors or measures” [27].

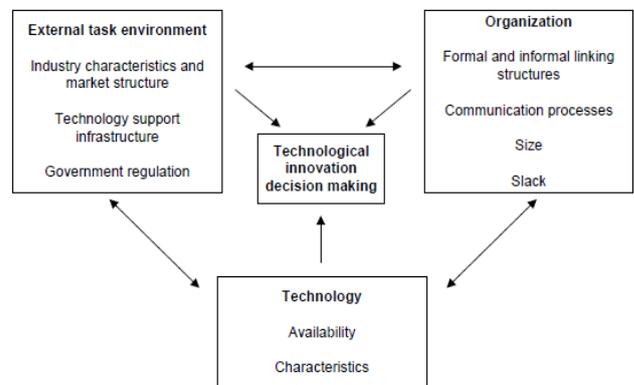


FIGURE I: THE TOE (TECHNOLOGY, ORGANISATION, AND ENVIRONMENT) FRAMEWORK [26]

The TOE technology aspect include all technologies that might be applicable to the enterprise including technologies currently being used by the enterprise, technologies that are available to the enterprise but are not being used and innovative technologies that might enable the enterprise to evolve and adapt [26] [27]. The organisational aspect of the TOE framework relates to all descriptive measures and resources of the enterprise (e.g. the number of employees and communication protocols) which effect may affect executive management decisions with regards to adoption and implementation [26]. Finally, the environmental aspect of the

TOE framework includes elements, such as the structure of the retail industry in SA, which might affect the adoption of technology within an enterprise [26].

### III. RESEARCH APPROACH AND DESIGN

The case study approach was chosen as an appropriate strategy to conduct this qualitative study. “A case study examines a phenomenon in its natural setting by employing multiple methods of data collection to gather information from one or a few entities (people, groups, or organisations)” [28]. The researchers actively solicited company documentation before and during interviews which were analysed to support the research.

The research was conducted within a leading African retailer which is part of a retail group with currently more than 4950 stores in 12 African countries. The retailer, used for the case study, currently has 2164 stores across Southern Africa and employs more than 15000 staff. The researchers chose the retailer’s head office, in the Western Cape region of SA, as the case site for this study. One of the researchers understands the company culture as he has been employed by the retail group for more than 7 years. Brands within the retail group are well-known household names in Southern Africa. The company has a strong customer focus and their core revenue comes from product sales through their stores to clientele.

The interviewees are listed in table I. All interviewees had degrees or diplomas. The gender split, 9 males and 3 females, is a fair reflection of the demographics found in the retail industry population. The sample population has vast IT retail experience (averaging 24 years) and had seen multiple IT strategies and technologies change over the past two decades.

TABLE I: RESEARCH POPULATION & SAMPLE

#	Position	Gender	Experience
P-1	Team Leader	Male	24 Years
P-2	Team Leader	Male	18 Years
P-3	Team Leader	Male	15 Years
P-4	Team Leader	Male	21 Years
P-5	DevOps	Male	10 Years
P-6	Enterprise Architect	Male	23 Years
P-7	Enterprise Architect	Female	25 Years
P-8	Director	Male	29 Years
P-9	Director	Female	26 Years
P-10	Director	Female	30 Years
P-11	Director	Male	32 Years
P-12	CIO	Male	35 Years

The interview protocol consisted of prepared structured (closed-ended) and unstructured (open-ended) questions. The researchers made notes during the interview of the interviewee’s comments, personal impressions and observations during the interview while audio recording each interview. The researchers used a thematic approach to analysing the documentation and qualitative data collected during the research using NVivo.

### IV. RESEARCH FINDINGS, ANALYSIS & DISCUSSION

The following three research questions drove the research:

- RQ1: What is the current understanding and perception of Digital Transformation within the SA retail industry?  
 RQ2: What digital technologies does the SA retail industry perceive are driving implemented to achieve Digital Transformation?  
 RQ3: What are the factors that influence the intended adoption of Digital Transformation within the organisation?

Twelve interviewees participated in the research strategy involving semi-structured interviews. Theme saturation point occurred from the ninth interview as no new themes emerged in the subsequent three interviews. Therefore, the researchers deemed the sample size sufficient (figure II).

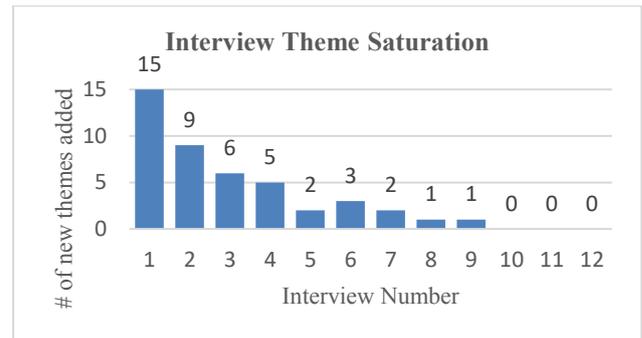


FIGURE II: INTERVIEW SATURATION PROGRESSION

#### A. Attitudes towards Digital Transformation

All interviewees demonstrated a positive attitude toward the adoption of digital transformation, highlighting common technologies that could assist the adoption of digital transformation within the enterprise. The majority of interviewees highlighted common factors that could influence the adoption of digital transformation and mentioned ways to overcome such factors. Furthermore, most interviewees indicated that implementing digital technologies will be beneficial for the enterprise leading to an increase in profit and reducing time to market, ultimately leading to an increase in market share.

#### B. Technology

A total of three factors were identified under the technology theme of which the two most significant were “perceived challenges” and “relative advantage”, refer to table 3 (the second column shows number of interviewees mentioning the theme). All participants perceived there to be both advantages and challenges by the adoption of digital transformation within the retail industry. The advantages and challenges could have a positive or negative effect on the adoption. The technology theme has been broken down into subthemes and will be discussed further in the next subsections.

TABLE II: TECHNOLOGY SUB-THEMES (NR OF MENTIONS)

Perceived Challenges	#
Infrastructure Impact	12
Security	10
Talent / Technical Skills	4
Relative Advantage	
Competitive Advantage	12
Reduced Cost	8

Time to Market	6
Customer Satisfaction	6
<b>Available Digital Technologies</b>	
G-Suite	12
Cloud	12
Artificial Intelligence	4
Machine Learning	2

*Perceived Challenges (+/-)*

A number of adoption barriers were highlighted in the literature review such as the lack of a clear and coherent digital strategy, talent, company culture, IT function transformation and Omni Channel retail capabilities. This was consistent with the responses from the participants with issues around security, workforce talent and resistance to change coming up the most.

It should be highlighted that despite digital strategy being highlighted during the literature review as one of the major challenges affecting the adoption of digital transformation, none of the participants mentioned this as a challenge.

**Infrastructure Impact (+)**

Several concerns were raised by participants around how the adoption of digital technologies, specifically cloud technology adoption, would impact their organisation’s infrastructure and thus would have a positive impact on digital transformation adoption. Most participants mentioned that the enterprise is currently busy with multiple projects to facilitate cloud adoption and that this will most certainly have a major impact on the existing infrastructure and data centres of the enterprise as P-3 stated “...we are in the process of digital transformation and part of it is moving from our own on premise infrastructure to hosted cloud-based infrastructure”. All participants were of the opinion that adopting cloud technologies will result in a saving for the enterprise as maintenance of the on-premise hardware will reduce significantly.

Another infrastructural concern that was raised was the need for high-speed low latency connectivity. P-6: “...with our cloud deployment as our need on high-speed low latency connectivity, for instance, is an infrastructure component that is important to improve on”.

**Security (-)**

Most participants highlighted that there is a security risk that must be considered with the adoption of cloud technologies as part of digital transformation: “You are moving outside the boundaries of your corporate network so that is a security risk.” (P-7)

P-7 is of the opinion that data is an organisation’s currency and the organisation should protect that intellectual property (IP) at all cost. P-8 also raised the concern for data protection by stating “...so by making us connected to everything at all times we have to make sure that we’ve got the bases in place to still protect a corporate organisation like we are”. P-2 further solidifies the security concern by stating “You’re ultimately putting everything into one place out there in the world, so security really needs to become a top priority”.

Further concerns were raised about SA’s Protection of Personal Information act (POPI) and payment card information (PCI) compliance. PP-2 states that “...the personal information act POPI requires information to be stored in a certain way with security applied. This actually becomes more important with cloud storage. Also, their payment card information also strict compliance rules that need to be adhered to...”. Although enterprises must take regulatory compliance into consideration it’s not necessarily a factor that would influence the choice of technology but rather the data generated by the technology as stated by P-11 “...I don’t think it necessarily influences the technologies that use but it certainly influences what you do with the data that that technology generates”.

**Talent - Technical Skills (+/-)**

As mentioned in the literature review, one of the most important critical success factors of digital transformation is to hire new digital talent to compliment or replace the existing workforce to ensure that the enterprise has the “right employees in the right place” [15].

Interviewees gave mixed responses when prompted to comment if the organisation possesses the necessary technical skills to implement digital technologies as part of their digital transformation strategy. Eight participants felt that technical skills within the organisation should not be an adoption barrier in today’s fast-changing IT environment: “...it’s not technical skills that will stand in our way” [P-8]. P-7 is of the opinion that the organisation could partner with an expert external third party to overcome any technology challenge by stating “We would have to partner with specialized partners in certain areas to take us on the journey. So if we partner with the right people I don’t think our technical skills will play such a big role”. However, P-2 felt that employees might fear new technologies and that new talent will be required to upskill existing employees by stating “Yes absolutely - the lack of technical skills as far as I can see brings about it a fear. New talent will be required for training of existing employees”.

*Relative Advantage (+)*

During the data analysis “Relative advantage” emerged as the second biggest factor under the technology major theme and having a positive impact on its adoption. Participants identified a number of perceived benefits (Table III), with customer satisfaction, competitive advantage and reduced cost being cited the most (second column = number of interviewees mentioning the theme).

TABLE III: PERCEIVED BENEFITS OF DIGITAL TRANSFORMATION

Customer Satisfaction	12
Competitive Advantage	9
Reduced Cost	7
Time to Market	6

All participants in the study highlighted that digital transformation would have a positive impact on customer satisfaction, ultimately providing the enterprise with a competitive advantage. Furthermore, by leveraging digital

technologies enterprises can collect data on customers and use data analysis techniques to offer customers specific goods and services as highlighted by P-12 “...*that if you start knowing your customer better you can offer him things that are very specific to him*”.

#### Available Digital Technologies (+)

Available digital technologies emerged as the third biggest factor under the under the technology major theme. An important factor pertaining to technology adoption is the availability of technological innovation [26]. It was important to identify the different digital technologies available as one of the objectives of this study was to identify potential use cases for digital transformation within the SA retail industry.

#### Google Suite (G Suite) (+)

G Suite is a set of Google applications that brings together essential services to help your business. This is a hosted service that lets businesses, schools, and institutions use a variety of Google products including Email, Google Docs, and Google Calendar. The adoption of Google Suite (G-Suite) technology, as a replacement of Microsoft Office, by the organisation was highlighted by multiple participants as an important step towards digital transformation. P-2 highlighted the benefit of multiple employees working on the same Google Doc from different locations while having a conference call by stating “*I mean it would it's nice to be able to sit and work on a document simultaneously with someone in Durban and Johannesburg while having a conference call*”.

#### Cloud (+)

The majority of participants stated that the use of cloud technologies plays a significant role towards the adoption of digital transformation. “*The cloud technology is one of the key technologies to enabled transformation. We are implementing Google cloud platform as our data lake option and we will leverage technologies such as AI in that platform*” [P-7]. The importance of providing the retailers and customers with close to real-time information on products and services were highlighted: “... *efficiently and close to real-time share information so that our feedback cycle from what we observe in our sales activity in the store can feedback all the way to our planning, manufacture, logistics and merchandising to close that loop so that that becomes more efficient*” [P-6].

#### Artificial Intelligence (AI) (+)

Artificial Intelligence (AI) refers to the ability of a digital computer, computer-controlled robot or systems to perform tasks commonly associated with intelligent beings like the ability to reason, discover meaning, generalize, or learn from past experience. The organisation will be leveraging their data in the cloud together with AI technologies to generate new valuable insights. P-7 states “*Put all of the data together in the cloud and then use the AI technologies to generate new insights for us. I think that there is a huge competitive advantage in machine learning and AI calls that technology will be able to derive insights way faster than a human will possibly be able to on volumes of data*”.

#### Machine Learning (MI) (+)

P-8 explained that the organisation already makes use of machine learning technologies to track if employee’s email has been infected by a virus. P-7’s personal view is that the organisation’s data in the cloud together with machine learning be of great benefit to the enterprise in the future.

#### C. Organisation

A total of nine factors were identified under the organisation theme during the semi-structured interviews. This made it the most populated major theme (Table IV).

TABLE IV: ORGANISATIONAL SUB-THEMES

Resistance to Change	12
Financial Resources	7
Big Data & Analytics	7
Digital Technology Readiness	6
Collaboration	5
Digital Strategy	4
Company Culture	4
Compatibility	3
Triability	3

The high number of factors identified by the participants suggested that organisational factors were dominant in their opinion when considering the adoption of digital transformation within the retail industry. The researchers will discuss these factors in the next subsections.

#### Resistance to Change (+/-)

The most cited factor under the organisation theme was resistance to change. Most participants highlighted that some employees have been working at the organisation for many years and is used to doing things a certain way: “...*we always has employees that are resistant to change so we might be leaving people behind if we don't put in a lot of effort to bring them on board and to take them with us on the journey. One of the factors is that people often say it's always been done in the same way that it's historically resistant to change*” [P-8]. P-6 highlighted the fact that before you can address resistance to change you must clearly define the scope of the project by stating “... *you need a clear definition of what your digital transformation is*”. Communication challenges can be addressed by a clear and well-defined scope which is clearly communicated to all employees. P-5 highlighted that as part of addressing the resistance to change within the organisation you may have to address issues with current processes. P-2 mentioned that change management is a very important factor in stating “...*change management is such a big thing. If you don't handle it correctly, they won't they won't adopt it, well not easily at least*”.

#### Digital Technology Readiness (+)

Some participants stated that the organisation needs to be prepared to adopt digital technologies. Preparing the organisation for digital technologies so that they are ready and willing to adopt new technologies will have a positive impact. P-7 stated that the organisation currently “...*rely heavily on the Gartner hype cycle, specifically the one for emerging technologies and retail technologies*” to ensure the

organisation is kept up to date with technology. Technologies that could present business value are proposed to the retailers to assess the retailer's appetite to adopt the technology. P-1 stated that each new technology is evaluated by a technical forum. The technology is assessed in a "sandbox" environment before testing it as a proof of concept (POC). Technologies are only approved for implementation after a rigorous evaluation process. P-5, who is part of the technical forum, states that each technology is assessed based on the organisation's requirement matrix and "...based on the requirements and our future requirements that we can think of we would select the technology that best fits our area and our vision".

#### *Financial Resources (-)*

Factors relating to financial resources in an organisation such as the cost of adopting new technologies or cost of changing existing technologies have a negative impact on whether an organisation will decide to proceed with implementing technology changes. Return on investment (ROI) plays a major part from the beginning when organisations decide on which technology projects will be implemented: "... if the return on investment is good enough then that's how I would motivate the use of new technology." [P-9]. ROI is very important in retail: "as low cost the retailer we are always striving to find more competitive ways to do things to be more cost-efficient" [P-9]. P-7 highlights the fact in some industries it is more difficult to secure funding for projects that in other by stating "...the retailers are very reluctant. In the banking sector and insurance sector money for major key investments is not such a big issue. In retail, you know the retailers are very reluctant to fork out the chequebook to buy these big investment items that are going to transform the companies".

#### *Big Data & Analytics (+)*

Data is being generated by a magnitude of sources within the organisation. The organisation harvests as much of the generated data as possible and store it in the cloud to use digital technology to generate new valuable insights for the business. P-7 stated that the organisation "...put[s] all of the data together in the cloud and then use[s] the AI technologies to generate new insights for us". The organisation benefited from the big data and analytics strategy in a number of ways. The organisation was "able to derive insights way faster than a human will possibly be able to on volumes of data" [P-7]. The sharing of information became easier: "...which then makes sharing information processes accessible to staff internally, external parties" [P-6]. Another benefit highlighted was that the adoption of cloud, analytics and the leveraging of big data gave the organisation the ability to better understand and know their customer. Marketing certain products to certain customers improved significantly as more information about the customer where collected, stored and analysed: "...if you can understand your customer you can better fulfil their needs and the only way to do that is with big data and data analysis" [P-2].

#### *Trialability (+)*

"Trialability is the degree to which an innovation may be experimented with on a limited basis..." [29]. A trial or test of the capabilities of digital technologies, in the form of a proof of concept (POC), is a way for an organisation to stifle any doubt or negative perceptions they may have about certain digital technologies and is a positive enabler of adoption. P-6 highlighted the fact that organisations don't always have to be the first to adopt new technologies, especially without proving that the adoption of new technologies will add value: "...you don't need to be the guinea pig and embark on new technology trends without proving that for yourself first. You don't want to go big bang on new stuff, you always want to take baby steps, always have POC's, test it out, monitor if you are achieving your goals and benefits you have set yourself and over time commit to more of that as you see that actually working within the organisation".

#### *Compatibility (+)*

"Compatibility is the degree of how consistent an innovation is perceived to be within an organisation and is affected by internal structures, strategy, values, experience and the needs of the business" [30]. P-7 confirmed that the retail organisation must investigate and demonstrate to the business that new digital technologies will add business value by stating that the organisation must "...determine what business value that is technologies will actually have for our retailers and also the appetite of the retailer to actually adopt that technology".

#### *Company Culture (+/-)*

Company culture can have a positive or negative effect on the adoption of digital transformation within the retail industry: "...if the culture is not ready or your culture is not very open to change, then that could be a barrier for you in terms of digital transformation" [P-7]. Experienced employees might see change as a risk to their careers. It is of utmost importance that the expectation of all employees are managed well, before, during and after embarking on a digital transformation journey as highlighted by P-6 "as people get a bit older, for them, it becomes a risk towards their career where the younger people are more eager to change. I think you need to balance that as well".

#### *Digital Strategy (+/-)*

The lack of a clear and coherent digital transformation strategy driven top-down by top management can be a negative factor adoption factor: "...you need a clear definition of what your digital transformation is. You do need to scope what you mean by that clearly. There is a risk that people might misunderstand what the context is and in terms of that there could be communication challenges" [P-6]. The need for a clear and coherent digital transformation strategy was emphasized by P-5: "...if traditional brick-and-mortar retailers want to survive they will need to transform and grow their e-commerce divisions." Their vision for digital transformation is clearly communicated to the entire organisation and is driven top-down by executives: "I think

the way the project is approached and communicated and the benefits explained to people will have a big impact on how positive the transformation will be accepted and how successful it will be” [P-9].

#### Collaboration (+)

Multiple participants highlighted that collaboration between employees within the organisation will have a positive impact on the adoption of digital transformation. P-8 stated that collaboration must be done between employees and clients to determine the scope for a project by saying “...collaborates with the clients at the highest level. Collaborate with what it is that they're trying to achieve”. The need to have communication tools available in an organisation with a distributed workforce was highlighted by P-2: “...we are a large group of people with the distributed management team across SA, there's definitely a need for communication tools that bring us together”.

#### D. Environment

Only three factors were identified under the environment theme (Table V). The retail customer was by far the most cited factor emphasizing its importance and customer-orientation.

TABLE V: ENVIRONMENTAL SUB-THEMES

Customer	12
Competition / Competitive Advantage	5
Time to Market	5
Connectivity	2

#### Customer (+)

The retailer customer was highlighted by most as one of the biggest driving factors of digital transformation within the SA retail industry, thus having a positive influence on adoption. The SA retail customer's behaviour when shopping and doing research before and during shopping is changing: “More and more customers are expecting digital transformation with buying online as well as doing online research and just the convenience of shopping anytime anywhere” [P-9]. Customers want to use their smart devices while shopping and expect information on products as and when needed: “...the customer has changed, the customer wants things and information at their fingertips, the customer wants to use new devices and in order to use their device of choice” [P-7].

Retail organisations must adapt and adopt digital technologies to ensure they cater for the needs of the changing customer by enriching the customer experience: [referring to the adoption of digital technologies to enrich the customer's shopping experience]: “...it needs to add customer value, it needs to improve the experience, it needs to reduce cost & risk, it needs to improve quality” [P-6].

#### a) Competition / Competitive Advantage (+)

Organisations have to change, adapt, transform and adopt digital technologies to stay competitive in the SA retail environment: “you will have to adapt and transform to stay competitive” [P-9]. A close eye is kept on what the competition is doing in the market, especially regarding digital technologies, to ensure sales and market share does not

decline: “...if our competitor is gaining market share or we are losing sales or whatever because of a competitor employing digital technologies it's definitely something that we will look at” [P-7]. Furthermore, the organisation is more reactive to externally visible digital technology innovations that could affect the customer's shopping behaviour: “When the customer's perception of innovation is positively affected by something, something that's clearly externally visible I do think that forces the organisation to adopt. You need to be more reactive in terms of an external visible technology influence than an internal one” [P-6].

#### Time to Market (+)

Time to market is a very important factor in a very competitive SA retail market. The perception arose during the semi-structured interview that digital technologies would improve time to market of products, thus having a positive impact on the adoption of digital transformation. Digital transformation coupled with improved business processes and activities will most definitely result in an reduced time to market and cost, increased market share and profit as mentioned by P-2 “...improve our processes and activities if we get this recipe right we will most definitely improve our time to market reduce our costs and increase our market share and hopefully this all leads to increase profits”.

#### Connectivity (-)

The retail group has a wide footprint of brick and mortar stores throughout Africa requiring an internet connection to trade and be operational. The lack of connectivity throughout Africa was raised as a factor that would negatively impact the adoption of digital transformation when P-3 stated “...I think connectivity is a massive issue given our big footprint and widespread brick-and-mortar stores connectivity is not consistently available everywhere”.

#### E. Technologies Driving Digital Transformation

The combined analysis from the literature review and semi-structured interviews revealed nine core technologies as driving the digital transformation within the SA retail industry. Table VI lists these, ordered by the combined number of participant responses per use case.

TABLE VI: TECHNOLOGIES DRIVING DIGITAL TRANSFORMATION

e-Commerce Solutions	12
Big Data & Analytics	9
Cloud	9
Artificial Intelligence	5
Bots	4
Machine Learning	3
Facial Recognition	3
Self-Checkout	2
RFID	2

The most cited technology across the board was to provide the retail group with an e-Commerce solution. All identified technologies were confirmed as relevant to the retail industry.

#### F. Summary of Findings

The findings can be used to answer the research questions.

### *Understanding and Perception of Digital Transformation*

The research study found that in terms of understanding digital transformation most participants had a general understanding of digital technologies that could be implemented as specific use cases to achieve digital transformation. However, none of the participants indicated that they are aware that digital transformation involves more than implementing certain digital technologies. The researchers found that there are a strong awareness and presence of digital technologies from participant responses received around the use of digital technologies within the SA retail industry which could have a positive or negative influence on adoption. Overall, perceptions from participants were mostly mixed to positive. The mixed perceptions were mostly due to date security concerns surrounding cloud adoption. Thus, the proposition is strongly supported.

### *Identified Technologies Driving Digital Transformation*

Participants identified nine technologies as driving digital transformation within the retail industry. Cloud technologies and data analytics were the most frequently cited and were the ones which the participants intended implementing. Specific benefits of these technologies included: satisfying changing customer needs, decreasing time to market, increasing customer value, which could be achieved by implementing digital transformation initiatives. Due to the agreement among and the specific technologies and benefits identified, this proposition is strongly supported.

### *Identified Adoption Barriers*

One of the goals of the study was to identify adoption factors, during the literature review and research strategy, and to provide the knowledge and information needed to the SA retail industry to ensure informed decision are made regarding potential future use. Several positive and negative different factors were identified across the Technology, Organisational and Environmental context.

Perceived challenges were highlighted as the biggest negative factor of the **Technological** theme, while relative advantage was highlighted as the biggest positive factor in the adoption of digital transformation. Data security when implementing cloud technologies was highlighted as a major concern / negative adoption factor while the adoption of cloud technologies and reduced onsite hardware was seen as a major benefit and stepping-stone to becoming digitally transformed.

The most adoption factors identified was under the **Organisational** major theme which indicated that participants thought about organisational factors. It became clear that resistance to change would be the biggest organisational barrier to overcome when undertaking digital transformation initiatives. Furthermore, it was highlighted that the cost of implementing digital technologies will play a big role and ultimately return on investment will be a deciding factor.

One of the most important findings that emerged from the **Environment** major theme was that the retail customer basically drives digital transformation. The changing customer, their needs and the change in how they shop was

highlighted by all participants. Thus customer-orientation rather than competitor analysis should drive the adoption of digital transformation in the SA retail sector.

## V. CONCLUSION, LIMITATIONS & FUTURE RESEARCH

The research study revealed that there was a good understanding of digital transformation from participants within the SA retail industry. However, this could be attributed to their level of training, years of retail experience, the position held within the organisation and the fact that the case site is currently busy with digital transformation projects. Therefore, this could not be generalized across the retail organisation or industry as a whole.

The core technologies perceived by participants as driving digital transformation in the retail industry with e-Commerce solutions, big data & analytics, and cloud adoption being mentioned the most. The findings from the study also affirmed that there is a consistent association between the adoption of new digital technologies and digital transformation. Finally, the findings identified 22 that affected the adoption of digital transformation within the SA retail sector grouped using the TOE model. **Technological** factors identified include the imperative to address technical challenges including securing a sound infrastructure (e.g. move some of the infrastructure into the cloud and secure a high-speed, low-latency connection ideally through a locally based cloud vendor); identified security risks include protection of IP as well as customer privacy and hire the correct technical skills. The relative advantages obtained include not only competitive advantage but also reduced costs, reduced time to market and improved customer satisfaction. Enterprises should make extensive use of digital technologies to create a customized personalized user experience through a comprehensive omni-channel approach. Digital technologies enable the enterprise to collect, analyze and interpret data which must be used to optimise their value chain, ultimately increasing profit.

The most prominent **organisational** issue identified was resistance to change, emphasizing the need for change management and a corporate culture embracing transformation and collaboration. Enterprises that add specialized digital change agents to their workforce to assist current employees through the digital transformation process will increase their success rate. Assessing the organisation's readiness, possibly through proof of concept sandbox testing is also important – these include aspects of the technology's trialability and organisational compatibility. Digitally mature enterprise must attract and recruit to ensure they don't have skill gaps. Reserving the appropriate financial resources is also important, as is ensuring that big data and effective data analytics are in place is also rated as a crucial step in digital transformation. A clear digital strategy which includes scope and objectives and is driven from the top down rounds off the organisational factors. The crucial **environmental** factor was, not unsurprisingly, a focussed customer-orientation for any technologies that are introduced i.e. does it help or add value for the customer as opposed to a knee-jerk competitor-driven

response. Many of these factors point towards the increased set of IT competencies that retail managers should master. Ultimately, enterprises must analyse the customer's behaviour through their entire shopping experience by making use of big data and analytics. An agile, creative, innovative workforce must elevate the customer's experience resulting in a competitive edge.

The extent to which these findings can be generalized to other countries depends on the structural and environmental similarity of the retail industry with the South African one which operates in a first world/third world context: infrastructure and scarce skills considerations may be more specific, but overall skills, readiness, corporate culture, change management, relative advantage and others are likely to be generalizable to many other country contexts.

Some limitations must be considered when interpreting the results of this study. The study provided a narrow focus on one large retail group, while the retail industry is made up of organisations of varying sizes. The seniority and the position that some of the interviewees fulfil in the case site enterprise resulted that they had limited time available for interviews. Also, the interviewee base was quite small but thematic saturation was reached after the ninth interview.

Future studies around the adoption of digital transformation in the retail sector could be conducted to identify new factors that might impact adoption. The research has identified the need for a study to be conducted across the SA retail industry in order to access a national view on digital transformation in different size retail organisations. In doing so the research will get a more in-depth view of positive and negative factors influencing the indented adoption of digital transformation and more potential use cases could be identified.

## VI. REFERENCES

- [1] E. Henriette, M. Feki and I. Boughzala, 2015. The shape of digital transformation: a systematic literature review. *MCIS 2015 Proceedings*, 431-443.
- [2] O. Gassmann, K. Frankenberger and M. Csik. The business model navigator: 55 models that will revolutionise your business. Cambridge: Pearson, 2014. <https://doi.org/10.3139/9783446437654.003>
- [3] S. K. Sia, C. Soh and P. Weill, 2016. How DBS Bank Pursued a Digital Business Strategy. *MIS Quarterly*, 15(2), 105-121.
- [4] L. Hudson and J. Ozanne, 1988. Alternative Ways of Seeking Knowledge in Consumer Research. *Journal of Consumer Research*, 14(4), 508-521. <https://doi.org/10.1086/209132>
- [5] G. Westerman, D. Bonnet and A. McAfee, 2014. *Leading digital: Turning technology into business transformation*. Massachusetts: Harvard Business Press, 13-15.
- [6] K. Dery, I. M. Sebastian, and N. van der Meulen, "The Digital Workplace is Key to Digital Innovation," *MIS Quarterly*, 16(2), 2017, 135-152.
- [7] I. M. Sebastian, J. W. Ross, C. Beath, M. Mocker, K. G. Moloney and N. O. Fonstad, 2017. How Big Old Companies Navigate Digital Transformation. *MIS Quarterly*, 197-213.
- [8] E. M. Rodgers, 1995. *Diffusion of innovations* (4th Ed.). New York, NY: Free Press.
- [9] G. C. Kane, D. Palmer, A. N. Phillips and D. Kiron, 2017. "Winning the digital war for talent." *MIT Sloan Management Review*, 58(2), 17-19.
- [10] H. Hansen and S. Sia, 2015. Hummel's Digital Transformation Toward Omni-channel Retailing: Key Lessons Learned. *MIS Quarterly*, 14(2), 132 - 149.
- [11] D. L. Rogers, 2016. *The digital transformation playbook: rethink your business for the digital age*, Columbia: Columbia University Press. <https://doi.org/10.7312/roge17544>
- [12] M. H. Ismail, M. Khater and M. Zaki, 2017. *Digital Business Transformation and Strategy: What Do We Know So Far?* University of Cambridge, Working Paper ITWeb IoT Survey. [http://v2.itweb.co.za/index.php?option=com\\_content&view=article&id=166391&Itemid=3087](http://v2.itweb.co.za/index.php?option=com_content&view=article&id=166391&Itemid=3087), [15 April 2017]
- [13] D. L. Soule, N. Carrier, D. Bonnet and G. F. Westerman, 2014. *Organizing for a Digital Future: Opportunities and Challenges*. MIT Center for Digital Business and Capgemini Consulting. Working Paper. <https://doi.org/10.2139/ssrn.2698379>
- [14] H. Gimpel, S. Hosseini, R. Huber, L. Probst, M. Röglinger and U. Faisst, *Structuring Digital Transformation: A Framework of Action Fields and its Application at ZEISS*. *Journal of Information Technology Theory and Application*, 19(1), 2018, 32-54.
- [15] T. Hess, C. Matt, A. Benlian and F. Wiesböck, 2016. Options for Formulating a Digital Transformation Strategy. *MIS Quarterly*, 15(2), 123-139.
- [16] H. Ghasemkhani, D. L. Soule and G. F. Westerman, *Competitive Advantage in a Digital World: Toward an Information-Based View of the Firm*. MIT Center of Digital Business, Working Paper, (2014). <https://doi.org/10.2139/ssrn.2698775>
- [17] M. Lane, A. Shrestha and O. Ali, 2017. *Managing the risks of data security and privacy in the cloud: a shared responsibility between the cloud service provider and the client organisation*. Seoul, The Bright Internet Global Summit.
- [18] M. Carroll, A. Van Der Merwe and P. Kotze, *Secure cloud computing: Benefits, risks and controls*. *Information Security South Africa (ISSA)*, IEEE. 2011, pp. 1-9. <https://doi.org/10.1109/ISSA.2011.6027519>
- [19] K. K. Patel, S. M. Patel and P.S.A Professor, 2016. *Internet of Things-IOT: definition, characteristics, architecture, enabling technologies, application & future challenges*. *International Journal of Engineering Science and Computing*, 6(5).
- [20] O. Vermesan and J. Bacquet, 2017. *Cognitive Hyperconnected Digital Transformation: Internet of Things Intelligence Evolution*, Eds. Demark: River Publishers. <https://doi.org/10.13052/rp-9788793609105>
- [21] O. A. Sawy, P. Kræmmergaard, H. Amsinck and A. L. Vinther, 2015. How LEGO Built the Foundations and Enterprise Capabilities for Digital Leadership. *MIS Quarterly*, 15(2).
- [22] A. Singh and T. Hess, 2017. How Chief Digital Officers Promote the Digital Transformation of their Companies. *MIS Quarterly*, 16(1).
- [23] N. Urbach, P. Drews, and J. Ross, 2017. Digital business transformation and the changing role of the IT Function. *comments on the special issue*. *MIS Quarterly*, 16(2).
- [24] E. Huré, K. Picot-Coupey and C. L. Ackermann, 2017. Understanding Omni-channel shopping value: A mixed-method study. *Journal of Retailing and Consumer Services*, 39, 314-330. <https://doi.org/10.1016/j.jretconser.2017.08.011>
- [25] P. C. Verhoef, P. K. Kannan and J. J. Inman, 2015. From multi-channel retailing to Omni-channel retailing: introduction to the special issue on multi-channel retailing. *Journal of retailing*, 91(2), 174-181. <https://doi.org/10.1016/j.jretai.2015.02.005>
- [26] R. DePietro, E. Wiarda, and M. Fleischer, *Processes of Technological Innovation*. Massachusetts, MA: Lexington Books, 1990.
- [27] J. Baker, *The technology-organization-environment framework Information systems theory*, New York, NY: Springer, 2012, pp. 231-245. [https://doi.org/10.1007/978-1-4419-6108-2\\_12](https://doi.org/10.1007/978-1-4419-6108-2_12)
- [28] I. Benbasat, D. K. Goldstein and M. Mead, "The case research strategy in studies of information systems," *MIS Quarterly*, 1987, 369-386. <https://doi.org/10.2307/248684>
- [29] M. Rodriguez, R. M. Peterson and H. Ajjan, 2015. CRM/social media technology: impact on customer orientation process and organizational sales performance. *Ideas in Marketing: Finding the New and Polishing the Old*, 636-638. Springer: Cham. [https://doi.org/10.1007/978-3-319-10951-0\\_233](https://doi.org/10.1007/978-3-319-10951-0_233)
- [30] A. Lin and N. C. Chen, 2012. Cloud computing as an innovation: Perception, attitude, and adoption. *International Journal of Information Management*, 32(6), 533-540. <https://doi.org/10.1016/j.ijinfomgt.2012.04.001>

# Aspects of Mobility of e-Marketing from Customer Perspective

Witold Chmielarz

University of Warsaw

Faculty of Management in Warsaw  
ul. Szurmowa 1/3

02-678 Warsaw, Poland

Email: witek@wz.uw.edu.pl

Marek Zborowski

University of Warsaw

Faculty of Management in Warsaw  
ul. Szurmowa 1/3

02-678 Warsaw, Poland

Email: mzbowski@wz.uw.edu.pl

Üyesi Mesut Atasever

Usak University

School of Applied Sciences in Turkey  
Ankara İzmir Yolu 8.Km Bir Eylül Kampüsü,  
Merkez, Turkey

Email: mesut.atasever@usak.edu.tr

**Abstract**—The main aim of this article is to identify students' opinions concerning the place, role and influence of electronic marketing tools on making purchases on the Internet. The authors have applied the division of e-marketing into its traditional and electronic forms, on desktop computers and mobile devices, which was significant due to diversified opinions of clients concerning its use. The studies have been carried out with the application of a CAWI method examining a convenient, partially randomly selected sample of clients (students) who are the most active in the Internet. The studies were aimed at evaluating specific e-marketing media and techniques which, in the customers' view, influenced shopping on the Internet. In particular, the respondents commented on the advantages, disadvantages and benefits resulting from the application of e-marketing on mobile devices. The conclusions and recommendations from the study may contribute to better use of these factors in order to facilitate consumers' purchases, not only in the Internet.

## I. INTRODUCTION

THE PRIMARY objective of this paper is to present the impact and significance of electronic marketing (e-marketing) in the purchasing process, based on the opinions of a selected group of potential clients. It is a next study conducted as part of a series of research analysing a similar group of respondents in the situation where the opinions on e-marketing are largely diversified, and a dynamic development of mobile devices may be observed. Simultaneously, it should be noted that the study is of supplementary nature in relation to comprehensive studies undertaken by the authors examining the quality of websites and mobile applications.

Electronic marketing is understood in this paper as a combination of all components related to information technologies, especially the Internet, in order to increase the willingness of potential customers to make purchases [13]. It is associated with many tools which are mainly applied on the Internet [18] as well as new sales and payment techniques [3]. It encompasses a wide range of themes connected with, for example, the evaluation of the possibilities of new devices (smartphones, tablets), users' response to new marketing forms and tools, the development of new e-marketing tools, etc. Kaznowski [10] believes that it is a significant part of the marketing strategy of an organisation. On the other hand, it is the result of a combination of modern marketing theories, use of information technologies [16] as well as the product

of project management, in particular, change and risk management [7]. In order to create a marketing strategy related to the promotion of products and services via the Internet, it is necessary to carry out a project consisting in, among others, building an e-shopping website and devising marketing tools which would help to promote this website on the Internet. For a marketing strategy to be successful, we should collect and examine the opinions of potential clients regarding the media and marketing techniques. If we consider all the above comments, then electronic market will represent all the above-mentioned marketing activities aimed at meeting operational, tactical and strategic goals with the application of the Internet infrastructure [5]. At present, mobile marketing is an essential part of electronic marketing. Due to the fact that it is perceived as all (advertising and promotional) activities using the functionalities of mobile devices [2], [9], it is difficult to distinguish between advertising available in browsers and special, dedicated smartphone mobile applications [8]. According to AMMA (The American Mobile Marketing Association), mobile marketing is understood as any form of marketing, advertising or promotional activity addressed to clients and transmitted via the mobile channel [15]. This definition of the phenomenon is the one applied in the present article. In addition, m-marketing offers basically unlimited possibilities of adapting the forms of promotion and communication to the needs of an individual recipient [12].

Electronic marketing was the object of many studies, both in the Polish and foreign markets, also from the point of view of a client [14], [6], [17], [19], [11], and new works analysing this field continue to appear. It is true that the majority of significant studies were published before 2015, the period of the most intense development of modern smartphones and tablets along with their dedicated applications, nevertheless, especially in mass surveys, new development trends and new phenomena appearing on this market are constantly being analyzed [20].

The authors of this article aim to distinguish some of the basic tendencies related to these new phenomena as well as implications for the future development of electronic marketing, including mobile marketing. The present studies, which main aim is to analyse the use of e-marketing among the users

of all kinds of computer devices used to access the Internet. The findings presented in this article, discussion and resultant conclusions constitute a report of the research involving a selected sample of Internet users in Poland at the beginning of 2019.

## II. THE ASSUMPTIONS OF THE METHODOLOGY AND PRESENTATION OF A STUDY SAMPLE

Following the previously conducted research [5], [4], the authors adopted the verified research procedure which consists of the following stages: constructing the first version of the survey questionnaire; verifying the questionnaire analysing the respondents' comprehension of the questions contained in the survey and the significance of the queries for the research, with the participation of randomly selected groups of respondents engaged in the pilot study, random selection of the groups of students for the study; making the verified and improved survey questionnaire available for the selected student groups (with the application of a CAWI – Computer Associated Web Interview method); analysis and discussion of the obtained findings; conclusions and possible directions for e-marketing development, on the basis of literature references and the authors' own studies.

In its final form, after eliminating the least significant questions and introducing changes aimed at clearer presentation of the remaining queries, the survey questionnaire included twenty-three substantive questions, divided into five groups and five questions related to the so-called demographics of the study sample. The scope of questions for the specified parts of the survey was as follows: electronic marketing environment, the effectiveness of electronic marketing, evaluation of e-marketing as a source of information on products/services, evaluation of the distinguished e-marketing media and techniques: the evaluation of selected e-marketing media, respondent's approach towards marketing on mobile devices and some demographics features.

The presented study was carried out in mid-March 2019. The research sample was selected as a partially convenient and partially random sample among the students of the University of Warsaw. An invitation to complete a survey questionnaire was distributed electronically among 356 students, both full-time and part-time courses, as randomly selected students' groups. 294 students completed survey questionnaires, which constitutes nearly 83% of the sample. This indicates nearly a threefold increase in the number of respondents from the same environment compared to the study of 2016 [5], which suggests increased interest in topics related to the possibilities of using the Internet for marketing purposes, especially in its mobile form. The selection of the sample consisting of students brought about certain limitations with regard to the possibility to interpret the findings. As the studies by Batorski and Płoszaj [1] indicate, the age group among which the studies have been carried out is a population which is most active in the Internet, most focused on innovation, and the one which is also the fastest to purchase and apply the latest technical solutions. Therefore, it is difficult to generalise the

obtained results to be indicative of the entire society. On the other hand, it is a group which for the above-mentioned reasons is the most competent to evaluate the tools used in the internet and mobile marketing, because they spend the greatest amount of time in the Internet, not only to obtain information, but also to make purchases and communicate with the shops, using websites and mobile applications many times a day.

In the analysed sample, more than 95% of the respondents were representatives of this most active social group. The group included individuals who were 18-24 years old, with the average age of slightly over 21, where all survey participants had secondary education. Among the respondents, there were 57% of women and 43% of men, which reflects the present gender structure of UW students. At present, in the examined study sample there were 55% of working students and 45% of students who were not professionally active. More than 52% of the respondents came from cities with over 500,000 residents, further 11% from towns with 100-500,000 inhabitants, 24% from the towns with 10-100,000 residents, less than 5% from small towns up to 100 inhabitants, and only 9% were from villages. In the present study, the share of students coming from large cities increased, mainly at the expense of people coming from rural areas.

## III. ANALYSIS OF THE OBTAINED FINDINGS AND THEIR DISCUSSION

The survey questionnaire was made available on the servers of the University of Warsaw. The questions were divided into several groups, and the analysis of the responses with the discussion and comments are presented below.

The first group of questions was of introductory nature. Its goal was to identify the conditions of using electronic marketing. The queries concerned the frequency of using the Internet, the type of most frequently visited websites, devices used for this purpose as well as the place and frequency of doing online shopping. The response to the first question appears to confirm the other findings [1] – all students use the Internet a few times a day. Undoubtedly, this was due to the popularity of mobile devices and – as it seems, a specific environmental culture of using them everywhere and at any time. This conclusion also results from the response to the following question, where over 23% of the respondents stated that it is the main and the only device which they use to connect with the Internet. Given that almost 12% of respondents use only a laptop and desktop computer for this purpose, this still confirms a clear advantage of this device over others. The greatest share of the sample – 44%, however, uses a combination of a laptop and a smartphone to connect with the Internet. As indicated in the comments section, the smartphone is mainly used to listen to music, communicate, obtain information and carry out small financial operations. Financial decisions which require careful consideration and extensive works or communications are usually associated with working on a laptop. In comparison with the situation from three years ago, the use of the smartphone as the only device to connect with the Internet declined (by nearly 10%),

and the use of smartphone and laptop increased (by almost 11%). The most frequently visited websites are social media websites (25%). Websites providing information/news are also popular - 12%. Thus, it emerges that the main and widely appreciated functions of the Internet are those which are associated with providing information or communicating. The use of search engines is also of primary importance with regard to providing information which the respondents require (18%). However, searching for a particular item is not always associated with purchasing it on the Internet: it is frequently only connected with looking for data concerning a given product or service. Nevertheless, e-shopping websites are most frequently visited by 19% of the respondents, and financial services by 21%. Thus, the area where electronic marketing might be applied is wide. The growing popularity of the use of mobile devices is demonstrated by the indicated places of accessing the Internet – nearly 93% of the respondents stated that they use it everywhere, and 13 times fewer people (7%) responded that they use it at work, at home or the university.

The second part of the survey concerned the perception of the effectiveness of the application of electronic marketing by internet users: their subjective evaluation of the phenomenon of e-marketing, its comparison with traditional marketing and evaluation of the potential advantage of e-marketing over traditional marketing. The respondents assess internet marketing as good or very good in over 84%, and only over 15% perceive it as satisfactory or non-satisfactory. This is probably caused by the opinion that 24% of the respondents are convinced that internet and mobile marketing is better, and over 50% believe that the greatest effectiveness is achieved through a combination of electronic and traditional marketing. In turn, nearly 23% of survey participants think that the two types of marketing are difficult to compare because they are addressed to different target groups. In the case of almost a quarter of recipients, there exists a belief that the effectiveness of marketing depends on the age of the recipients and the most frequent use of media (smartphone versus laptop) associated with it. The respondents regard continuous availability (33%, 25% in 2016) via mobile and remote or desktop devices and the possibility of buying items after clicking on the advertising field (via link) (nearly 26% as compared to 19% in 2016) to be the greatest sources of advantage of marketing in the Internet over traditional marketing. Also, the previously emphasised possibility to obtain more information about a product or service (25%) is of considerable importance.

The third part of the survey concerned issues related to the use of sources of obtaining information on products and services on the Internet and outside the Internet as well as its subsequent application. Among the analysed sample, the Internet proved to be an a decisively dominant medium (86% individuals) to access information about products and services (as compared to previous score at the level of 33%). In combination with the information obtained from a circle of friends and colleagues, this comprises over 98% of the places of obtaining commercial information. The importance of such media as television, radio, press, leaflets or paper

information materials appears to be nearly non-existent, which points to little interest in this form of marketing among the representatives of this social group. Comparison engines turned out to be the most common tool (as it is believed by over 40% of the respondents) to search for information about products and services on the Internet. On the one hand, individuals are eager to use them, on the other hand, they do not perceive them as a tool which would be of crucial importance from the point of view of the effectiveness of e-marketing. The second place is taken by social media (nearly 31% share in the respondents' opinions). Even three years ago such an opinion would be encountered with disbelief; however, at present the influence of social media is becoming more and more important. This is also evidenced by the high, third place of blogs (15%). It is important to point out that even though a blog is in fact seen as a source of largely subjective information, it is still a medium which shapes consumers' tastes and views in certain sectors (e.g. fashion and cuisine). E-marketing offering different advertising forms contained on websites and carried out via emails is losing its importance (in total the score amounts to less than 9%). This form of advertising, which until recently was seen as dominant in this type of marketing, in a sense, is already regarded as a traditional form.

In the fourth part of the survey the respondents were asked to evaluate particular media and e-marketing techniques: the effectiveness of the applications of the media, approach towards selected e-marketing techniques, places in the ranking of products which induce consumers to make purchases, elements which respondents pay particular attention to and those which attract them the most as well as the evaluation of the respondent's approach to placing particular elements of e-marketing in marketing media and on various types of devices. The evaluation of the effectiveness of selected electronic marketing media was based on a four-point scale from: unsatisfactory, satisfactory, good and very good. The highest rated techniques were: presence in social media (26% of very good scores), clarity and attractiveness of a website (24% of very good scores) as well as the presence in mobile solutions (20%). The highest number of good scores were obtained by positioning (18%) and sponsored links (15%). In the latter case, the opinions were divided because slightly more people (18%) evaluated them only at a satisfactory level. Banners, links to other websites (19% and 18% respectively) and newsletters (21%) were evaluated at the border between satisfactory and unsatisfactory. The most unsatisfactory technique was related to advertising mailing messages (43% of unsatisfactory scores). This survey section was also aimed at creating a specific ranking of factors which motivate clients to make purchases. In this ranking, the first place among the responses was taken by the clarity and attractiveness of a website (33% of views). The second position was occupied by the presence in social media (22% of opinions). The subsequent places were taken by factors such as discounts after exceeding a specific value of the purchase (17% responses) and positioning (11%). The last positions in the assessment were taken by pop-up

windows (43% in the last position), e-mailing advertisements (20% in the penultimate position). According to respondents' opinions, clients pay the most attention to graphic elements (34%) as well as the innovativeness and attractiveness of the presentation (29%). They pay the least attention to technical elements of e-marketing such as: text (8% of the surveyed students believe it is the case) or the sound and music (14% of responses). So, what would attract them to visit the website? In the views of the study participants, at present, the most efficient in this regard are elements such as short videos (28%) and large graphic banners between a logo and the content (16%). The least effective are: buttons (5%) and pop-up windows (8%). The above ranking shows a specific transition of the existing clients to more modern technical elements of e-marketing and "fatigue" with the forms which are frequently encountered in the current practice of using the Internet. From the point of view of a client, the greatest acceptance for placing e-marketing in selected marketing media was recorded in the case of e-shopping websites (23% of responses), social media websites (20%) as well as company and news websites (15% and 16% respectively).

The last group of survey questions concerned the respondents' approach to the phenomenon of marketing on mobile devices, namely: advantages and disadvantages of m-marketing, benefits of m-marketing and the effectiveness of m-marketing techniques in relation to the client. Among the greatest advantages of m-marketing, the respondents mainly distinguished the fact that it is available at all times and everywhere (24%) and it can apply a personalised advertising message (21%). The last positions were taken by the high effectiveness of this medium (8%) as well as the fact that it can be treated as a determinant of modernity (10%). The advantages of m-marketing bring direct benefits for the client. Among the selected benefits, the most important factor (37% of the responses) was the use of NFC technique or QR codes (e.g. train tickets). According to the survey participants, the second significant benefit was geolocation and mobile navigation (31%). The subsequent positions were taken by the possibility to create mobile websites (13%) and SMS marketing (9%). The biggest disadvantage of m-marketing is the necessity of longer screen scrolling (34% of respondents believe it is the case) and increasing difficulty of getting rid of advertising messages (33%). Another negative factor is the fact that they take too much space on a screen which is already rather small (20%). The smallest number of people believe that advertising on mobile devices is too general and the graphic presentation is of lower quality (3-4%).

According to the survey participants, the greatest influence is indicated in the case of graphic advertising elements (66%). The next place is taken by the video advertising of the application (17%) and graphic advertising of the application (11%). The remaining kinds of mobile advertising are of limited importance, namely, they constitute only 5%. In the last six months, the aspects which had the greatest impact on respondents' purchases included: the use of mobile applications (44%), using geolocation and mobile navigation

(17%) as well as SMS marketing (14%). The remaining m-marketing techniques did not exert any significant influence on the purchases made by the respondents in the last six months.

#### IV. CONCLUSIONS

The conducted and presented studies lead to the following conclusions:

- the examined population is "immersed" in the Internet nearly all the time, using mainly mobile devices to search information, exchange communications, enjoy broadly defined entertainment (music, films, computer games), as well as make purchases or carry out financial transactions. This tendency has strengthened in the last three years,
- the opinion about electronic marketing and its impact on purchases is still very high. This is not reflected in the value of purchases; nevertheless, this results from appreciating the informative function of the Internet. Continuous availability and convenience of the use is not only or mainly associated with making purchases, but it also serves to obtain information about a product or service. The decisions concerning the purchase and the way this operation is being carried out (via the Internet or traditionally) are taken later,
- the attitude of the respondents towards comparison engines is unclear. On the one hand, nearly everyone uses them; on the other, clients do not perceive them as the most important tool which might be seen as a specific advantage of e-marketing over traditional marketing,
- the effectiveness of e-marketing media, in the respondents' opinion, depends mainly on the presence in the social media and characteristic features of the website (its clarity and attractiveness); in the ranking of the factors inducing consumers to make purchases, apart from the above aspects, the respondents list also discounts offered after exceeding a certain amount of money,
- pop-up windows and spam mailing are the two most disliked elements in e-marketing,
- the respondents mainly pay attention to such technical elements of e-marketing, like graphic elements, in particular, short videos, appearing mainly on social media websites,
- the irritation associated with excessive advertising in mailings is growing; while the degree of acceptance of e-marketing received via traditional and modern devices is rather high (28%). The studies concerning this very phenomenon in relation to websites [e.g.] show that this solution appears to be the most undesirable with regard to the evaluation of the website quality. The greatest level of acceptance for video marketing on mobile devices undoubtedly also plays an important role in this respect,
- we may also observe a phenomenon of a specific shift of the interaction with the Internet from traditional to mobile devices and more and more common blurring of the boundaries between mobile laptops and tablets due to the greater universal use of laptops. The dominating position of smartphones in everyday life also has more

and more influence on the evaluation of e-marketing in its mobile form,

- the advantages of m-marketing result from its continuous availability and a possibility to personalise the message; the disadvantages mainly consist in the fact that it occupies a large part of the screen and its related necessity of longer scrolling or the fact that such an advertisement is more and more difficult to remove from the screen.

The limitation of the study was the fact that it was carried out among a rather uniform sample of respondents coming from academic environment. As previously mentioned, this was the most active group with regard to new technologies, and the obtained findings tend to present a somewhat idealised view of the clients' relation towards both the technologies themselves as well as the operating media of electronic advertising. The study should be extended to include also other social groups which do not use the Internet to such an extent, both in their private life and economic activity. This would allow for a more comprehensive, holistic view of the possibilities of e-marketing applications. On the other hand, international and intercultural studies seem to be a very interesting direction for further studies, which would allow for specific universalisation of the obtained findings.

#### REFERENCES

- [1] D. Batorski, A. Płoszaj, *Diagnoza i rekomendacje w obszarze kompetencji cyfrowych społeczeństwa i przeciwdziałania wykluczeniu cyfrowemu w kontekście zaprogramowania wsparcia w latach 2014-2020*, Warszawa, 2012, [http://www.euroreg.uw.edu.pl/dane/web\\_euroreg\\_publications\\_files/3513/ekspertyza\\_mrr\\_kompetencjegyfrowe\\_2014-2020.pdf](http://www.euroreg.uw.edu.pl/dane/web_euroreg_publications_files/3513/ekspertyza_mrr_kompetencjegyfrowe_2014-2020.pdf).
- [2] D. Bernauer, *Mobile Internet - Grundlagen, Erfolgsfaktoren und Praxisbeispiele*. Vdm Verlag Dr. Müller, 2008.
- [3] *Charakterystyka bankowości elektronicznej*, ed. A. Gospodarowicz, *Bankowość elektroniczna. Istota i innowacje*, Warszawa, Wydawnictwo C.H. Beck, 2018.
- [4] W. Chmielarz, *Study of Smartphones Usage from the Customer's Point of View*, *Procedia Computer Science*, Elsevier, Vol. 65, 2015, pp. 1085-1094. DOI: 10.1016/j.procs.2015.09.045
- [5] W. Chmielarz, M. Zborowski, *Aspects of mobility in e-marketing from the perspective of a customer*. eds. M. Ganzha, L. Maciaszek & M. Paprzycki, *Proceedings of the 2016 Federated Conference on Computer Science and Information Systems* [online]. Warsaw, Polskie Towarzystwo Informatyczne, 2016, pp. 1329-1333. DOI: 10.15439/2016F112
- [6] T. Gao, F. Sultan, A. J. Rohm, *Factors influencing Chinese youth consumers' acceptance of mobile marketing*, *Journal of Consumer Marketing* 27/7, 2010, pp. 574-583. DOI: 10.1108/07363761011086326
- [7] J. Hasan, *Analysis of E-marketing Strategies*, *Studia commercialia Bratislavensia*, Volume 4; Number 14 (2/2011), 2011, pp. 201-208. DOI: 10.2478/v10151-011-0006-z
- [8] N. Hatalska, 2016, <http://hatalska.com/slangoskop/marketing-mobilny/>. DOI: 10.2478/v10151-011-0007-y
- [9] D. Hovancakova, *Mobile Marketing*, *Studia commercialia Bratislavensia*, Volume 4; Number 14 (2/2011); 2011, pp. 211-225.
- [10] D. Kaznowski, *Nowy marketing w Internecie*, Difin, Warszawa, 2007.
- [11] M. Kiba-Janiak, *The Use of Mobile Phones by Customers in Retail Stores: a Case of Poland*, *Economics & Sociology*, Vol. 7, No 1, 2014, pp. 116-130. DOI: 10.14254/2071-789X.2014/7-1/11
- [12] S. Konkol, *Marketing mobilny*, Helion, Gliwice, 2010.
- [13] X. Meng, *Developing Model of E-commerce E-marketing*, *Proceedings of the 2009 International Symposium on Information Processing (ISIP'09)*, Huangshan, P. R. China, August 21-23, 2009, pp. 225-228.
- [14] G. Roach, *Consumer perceptions of mobile phone marketing a direct marketing innovation*, *Direct Marketing An International Journal* Vol. 3 No. 2, 2009, pp. 124-138. DOI: 10.1108/17505930910964786
- [15] J. Salo, J. Sinisalo, H. Karjaluto, *Intentionally developed business network for mobile marketing: a case study from Finland*, *Journal of Business & Industrial Marketing* 23/7, 2008, pp. 497-506. DOI: 10.1108/08858620810901257
- [16] S. Sun, *Innovation Mode and Strategy Research on Small and Medium-sized Enterprise E-marketing in Post Financing Crisis*, *Contemporary Logistics* 04, 2011, p. 13. DOI: 10.5503/J.CL.2011.04.003
- [17] U. Świerczyńska-Kaczor, *e-Marketing przedsiębiorstwa w społeczności wirtualnej*, Difin, Warszawa, 2012.
- [18] A. Sznajder, *Technologie mobilne w marketingu*, Wolters Kluwer S.A., Warszawa, 2014.
- [19] J. Wielki, *Modele wpływu przestrzeni elektronicznej na organizację gospodarcze*, Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław, 2012.
- [20] *Wirtualnemedi*, 2018, <https://www.wirtualnemedi.pl/artykul/najpopularniejsze-serwisy-internetowe-i-aplikacje-mobilne-w-polsce-w-listopadzie-2018-r-dane-gemius-pbi-z-listopada>.



# Network Effects in Online Marketplaces: The Case of Kiva

Haim Mendelson and Yuanyuan Shen

*Abstract*—Advanced information technologies have enabled the development of online marketplaces that connect businesses and people on a global scale. Much of the analysis of the adoption, growth and engagement on these marketplaces in the extant literature is based on the premise that they are characterized by network effects—a premise that has major implications for their deployment, implementation and management. In this paper we test this premise using data from Kiva, the world’s largest online, peer-to-peer social lending marketplace. We find that while network effects are strong and significant during the early growth phase of the marketplace, they become weak or disappear once the marketplace stabilizes.

*Keywords*—Online marketplaces, network effects, peer-to-peer lending, online services.

## I. INTRODUCTION

ADVANCED information technologies are changing the structure of economic activity, with many traditional processes being transformed through the use of electronic marketplaces. Activities such as buying, selling and lending are moving from the established but labor-intensive and inefficient brick-and-mortar format to online marketplaces that increase efficiency, transparency and effectiveness and are already a major sector of the economy. The research questions addressed in this paper are: (i) are online marketplaces characterized by network effects? (ii) How does the answer depend on the growth phase of the marketplace? We address these questions using data from Kiva, the world’s largest social lending marketplace.

Network effects, also referred to as network externalities, reflect a positive relationship between the installed base of users on a platform and its value to users [1],[2],[3]. They are *direct* when there is a direct positive relationship between the size of the installed base and the value to users within that installed base. The classic example is the telephone network: adding a new user to the network increases the number of potential calls users can make, which increases the utility users derive from the network [4]. *Indirect* network effects arise when (i) the network is based on two complementary components, say  $A$  and  $B$ ; (ii) there is a positive relationship between the installed base of  $B$  and the value to users of  $A$ , and (iii) there is a corresponding positive relationship between the installed base of  $A$  and the value to users of  $B$  [1]. This results in a positive feedback loop between the installed bases of  $A$  and  $B$ : an increase in the installed base of  $A$  makes the

network more attractive to the  $B$ s, and as more  $B$ s join the network, it becomes more attractive to the  $A$ s. This means that more  $A$ s attract yet more  $A$ s indirectly through the  $B$ s—hence the term *indirect*. In this paper we test the existence of this positive feedback loop.

Network effects have a major impact on the way technology-based solutions are deployed and managed as they affect choices of efficiency, effectiveness and speed: in the presence of network effects, a highly-efficient and effective solution that does not achieve critical mass may fail regardless of its technical or economic merit. Further, if network effects are sustainable, a solution that manages to control a large user base may prevail even when it is inferior on a stand-alone basis [1]. Thus, network effects have a paramount impact on the deployment and management of platforms, and in particular—on online marketplaces.

In a peer-to-peer online lending marketplace, prospective borrowers post loan requests online either directly, on their own, or indirectly, through marketplace partners. Lenders browse the loan requests and decide which loans they would bid on. Lenders who wish to fund a loan submit conditional or unconditional funding commitments to the marketplace. The marketplace then matches loan requests to funding commitments, funds some of the loans, and services them until they are repaid (or until they default). It is commonly assumed that such lending marketplaces are characterized by indirect network effects between lenders and borrowers, as more lenders increase the probability of a loan request being funded, and more borrowers make the market more attractive to lenders, who can better diversify their loans and are more likely to find a match they are willing to fund. The latter consideration is important on Kiva, where lenders seek to support entrepreneurs with particular characteristics, and with more entrepreneurs and loan requests on the site, a lender is more likely to find one she is willing to support.

Network effects were found in a variety of industries ranging from telecommunications to Information Technology (cf. [5], [6], [7]). However, while the theoretical literature views network effects as an inherent feature of online marketplaces, we could not identify an empirical study that directly confirmed their existence in online peer-to-peer lending marketplaces. In this paper, we narrow this gap by investigating whether network effects actually exist on Kiva and how they depend on the growth phase of the marketplace.

The rest of the paper is organized as follows. Section II is a Kiva overview. Section III outlines our research hypotheses. Section IV describes our data and test methodology. Section V presents our results. We briefly conclude in Section VI.

H. Mendelson (haim@stanford.edu) is with the Graduate School of Business, Stanford University, Stanford, CA 94305, USA (see <https://www.gsb.stanford.edu/faculty-research/faculty/haim-mendelson>). Yuanyuan Shen (anashen@alumni.stanford.edu) is with the Graduate School of Business, Stanford University, Stanford, CA 94305, USA.

## II. KIVA

Founded in October 2005, Kiva operates a website where entrepreneurs from developing countries post loans through field partners—microfinance institutions, social businesses, schools, and other non-profit organizations. Loans come from individual lenders from across the globe, primarily from developed countries. Between October 2005 and June 2019, Kiva funded \$1.32 billion in loans extended to 3.3 million borrowers from 1.8 million lenders. These loans had an impressive repayment rate approaching 97%.

Each month, Kiva's field partners post on the Kiva website loan requests on behalf of the entrepreneurs they represent. Loan terms average 1.5 years. Lenders browse the loan requests and may contribute \$25 or more to fund the loans they select.

Loan requests remain posted on Kiva for up to 30 days. If a loan is not fully funded within that period, it expires and all lenders' commitments are refunded. If the loan is fully funded, Kiva's field partner sends the money to the entrepreneur. As of June 2016, about 95% of loan requests were fully funded. As the borrower repays the loan, the field partner returns the principal through Kiva to the lenders who funded it.

Kiva lenders are social investors who receive no interest and make no profit on their loans (however, Kiva's field partners may charge interest on the loans they make). Loans serve the needs of poor, under-served, or financially excluded (e.g., unbanked or underbanked) populations and aim to achieve a social or environmental impact.

We obtained our data from Kiva's data snapshot on build.kiva.org, augmented by querying Kiva's API. In addition to basic data on borrowers and prospective lenders, we have loan-specific information through June 2016. Our sample period is January 2007 to June 2016. In our sample, the majority (84.8%) of loan requests come from Asia, Africa and South America while most (89.4%) lenders come from developed countries in North America and Europe. As a macroeconomic control, our regressions use the effective yield on the ICE BofAML Emerging Markets Corporate Plus Index [8], which is available on a daily basis and is expected to influence loan funding in emerging markets (we also used the GDP growth rate and unemployment rate, which turned out insignificant).

## III. RESEARCH HYPOTHESES

In this paper, we formulate two key hypotheses on potential network effects and then test them using data from Kiva. As discussed above, the theory of network effects implies a positive feedback loop between the number of lenders and the amount of open loans (measured by their number or aggregate dollar amount): with more lenders, the platform should attract more loan requests, and with more loan requests, the platform should attract more lenders. This results in two key hypotheses:

*Hypothesis 1:* The dollar amount and number of open loans on Kiva should increase in the lagged number of active lenders.

*Hypothesis 2:* The number of active lenders on Kiva should increase in the lagged number and amount of open loan requests.

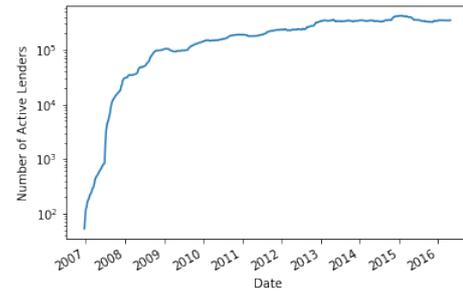


Fig. 1. Number of active lenders on Kiva

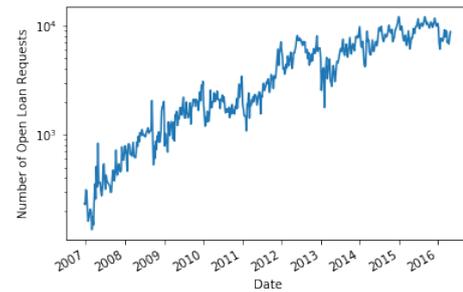


Fig. 2. Number of loan requests open on Kiva

Network effects and the associated feedback loop require *both* Hypotheses to hold.

## IV. DATA AND METHODOLOGY

We construct weekly time series for Kiva's performance over our sample period (2007 - 2016). We identify two phases in the development of the Kiva platform: a *growth* phase covering the first half of our sample period, and a more stable phase during the second half of our sample period. During the growth phase, both the amount of loans requested and the amount funded grow quickly. Then, the growth in both supply and demand flatten out. Further, during the second subperiod, the gap between the number of loans requested and the number of loans funded becomes larger.

Figure 1 shows the number of active lenders (i.e., those who have bid on at least one loan in the past six months) and Figure 2 shows the number of open loan requests on Kiva each week over our sample period, both on a logarithmic scale. The figures show how growth has abated between the earlier growth period and the latter stable period, when the number of active lenders flattens out and the growth in open loan requests deteriorates.

These patterns show distinct differences between the earlier growth period and the latter stability period. While the empirical network effects literature typically finds them at the early growth phases of new technologies, it does not recognize the stark differences between the early growth and stability periods. Madden and Dalzell [9] study the early growth of mobile telephony and attribute differences between high- and low-income countries to non-linear network effects.

We argue that network effects should be more prominent during the earlier growth period, when the growing installed

base is a key driver of adoption. Once the platform stabilizes, we expect other factors and tactical moves undertaken by the platform (e.g., field partner selection, budgeted loan amounts, etc.) to overtake the network effects as drivers of performance. Indeed, while peer-to-peer lending has been growing rapidly around the globe [10], Kiva has experienced a declining loan fulfillment rate, and during the 2011-12 period it started honing its business model, actively managing demand and supply down to the level of the types of loans made [11], [12]. In addition, Kiva was facing increased competition from new social microfunding sites (e.g., MyC4 in Europe, Wokai in China and MicroPlace in the U.S.), as well as from for-profit lending sites that catered to entrepreneurs (e.g, Zopa in Europe and CreditEase in China; these sites, which charge interest, competed with Kiva since many field partners charge interest to the entrepreneurs they serve). In addition, Feldman et al. show theoretically that the performance of peer-to-peer systems can degrade significantly as a result of user turnover [13]. We thus expect our network effects hypotheses 1 to 2 to hold during the early growth period and to become substantially weaker or altogether disappear during the latter stability period.

To test for these different behaviors, we divide our observations into two halves, (i) January 2007 to August 2011 and (ii) September 2011 to June 2016. Table I displays the correlation matrix among our key variables for the two sample subperiods.

TABLE I  
CORRELATIONS BETWEEN THE NUMBER OF ACTIVE LENDERS, THE DOLLAR AMOUNT AND THE NUMBER OF LOAN REQUESTS

Variable	Amount of Loan Requests	No. Loan Requests
Jan 2007 - Aug 2011		
No. Active Lenders	0.87	0.87
Amount of Loan Requests	-	0.99
Sept 2011 - June 2016		
No. Active Lenders	0.46	0.46
Amount of Loan Requests	-	0.95

In Table I, we observe a strong correlation between the number of active lenders and the number and dollar amount of loan requests during the first subperiod. These correlations substantially decline in the second subperiod. Due to the correlations among our key explanatory variables, we test each of our hypotheses separately. For Week  $t$ , we denote by  $n_t$  the number of active lenders (defined as those who placed at least one bid over the past six months), by  $a_t$  the dollar amount of open loan requests, and by  $l_t$  the number of open loan requests. We estimate (heteroskedasticity-corrected) OLS regressions using the specification below to test Hypothesis 1 that the dollar amount of loan requests increases in the number of active lenders:

$$\log(a_t + 1) = \alpha_0 + \alpha_1 \log(n_{t-1} + 1) + \vec{\alpha}_3 \vec{x}_{t-1} + \epsilon_t, \quad (1)$$

where the vector  $\vec{x}_{t-1}$  comprises the default rate on Kiva loans as of the end of Week  $t - 1$ , the effective yield on the ICE BofAML Emerging Markets Corporate Plus Index [8], dummy variables representing the year of  $t$ , and a trend variable.

An alternative test of Hypothesis 1 focuses on the relationship between the number of loan requests and the number of

active lenders:

$$\log(l_t + 1) = \beta_0 + \beta_1 \log(n_{t-1} + 1) + \vec{\beta}_3 \vec{x}_{t-1} + \zeta_t. \quad (2)$$

To test Hypothesis 2 that the number of active lenders increases in the dollar amount of loan requests, we estimate the regression:

$$\log(n_t + 1) = \gamma_0 + \gamma_1 \log(a_{t-1} + 1) + \vec{\gamma}_3 \vec{x}_{t-1} + \omega_t. \quad (3)$$

Likewise, we test whether the number of active lenders increases in the number of open loan requests by estimating the equation:

$$\log(n_t + 1) = \kappa_0 + \kappa_1 \log(l_{t-1} + 1) + \vec{\kappa}_3 \vec{x}_{t-1} + \psi_t. \quad (4)$$

In Equations (1) through (4),  $\epsilon_t, \zeta_t, \omega_t, \eta_t$  are random noise.

## V. RESULTS

Table II shows the results of our OLS estimations using White's method to account for heteroskedasticity [14].

As hypothesized, we observe positive and strongly-significant coefficients for our network effect variables during Kiva's growth period: a 1% increase in the number of active lenders results in a 0.31% increase in the dollar amount and a 0.46% increase in the number of open loan requests. A 1% increase in the dollar amount of open loan requests leads to a 0.17% increase in the number of active lenders, and a 1% increase in the number of loan requests leads to a 0.32% increase in the number of active lenders. These results are both economically meaningful and strongly statistically significant, confirming our hypotheses. Given that *both* hypotheses hold, we have the feedback loop confirming the existence of network effects during the growth period.

For the second subperiod (September 2011 through June 2016), the network effect coefficients become small and, for the most part, insignificant (the exception is the coefficient of the number of open loan requests in Eq. (4), which is significant at the 90% level). Based on equations (1)-(3), *both* hypotheses are rejected. Because a rejection of either one of our hypotheses leads to the rejection of network effects, we conclude that there are no meaningful network effects during the second subperiod. These results are consistent with our argument for the differences between the two subperiods (Section IV).

Equations (3)-(4) use the number of active lenders to estimate the dependent variable, which leads to autocorrelated residuals. To address this issue, we reestimated equations (3)-(4) by differencing the number of active lenders, using  $(n_t - n_{t-1})$  as our dependent variable. The results are shown in columns (3a) and (4a) of Table II. The main difference between the two specifications is that in equation (3a), the significance of the amount of loan requests declines in the first subperiod and increases in the second. Our conclusion remains intact: there are strong and significant network effects in the first subperiod whereas in the second subperiod, there are no network effects, as Hypothesis 1 fails to hold.

TABLE II  
RESULTS FOR HYPOTHESES 1 AND 2

Regressors	Jan 2007 - Aug 2011						Sept 2011 - June 2016					
	Eq. (1)	Eq. (2)	Eq. (3)	Eq. (3a)	Eq.(4)	Eq. (4a)	Eq. (1)	Eq. (2)	Eq. (3)	Eq. (3a)	Eq.(4)	Eq.(4a)
Intercept	9.67*** (0.89)	1.52+ (0.79)	8.03*** (0.71)	-10.432+ (6,118)	8.21*** (3.87)	-7.040 (3,046)	11.93** (3.79)	4.22 (0.15)	12.48*** (17,294)	-35.236* (0.09)	12.55*** (10,908)	-18.462+ (1,299)
No. Active Lenders	0.31*** (0.09)	0.46*** (0.08)	-	-	-	-	0.13 (0.31)	0.22 (0.30)	-	-	-	-
Amount Loan Requests	-	-	0.17** (0.05)	866.83+ (470.40)	-	-	-	-	0.02 (0.01)	2.485* (1,171)	-	-
No. Loan Requests	-	-	-	-	0.32*** (0.06)	1,248** (482)	-	-	-	-	0.02+ (0.01)	2.446+ (1,299)
Default Rate	-1.67 (0.79)	-2.23 (6.03)	-3.34 (2.30)	5.578 (41,893)	-3.88 (2.49)	4,542 (42,179)	102.65*** (18.71)	96.22*** (16.57)	6.50* (3.11)	-26.047 (263,670)	6.08+ (3.11)	-12.015 (268,525)
Interest Rate	-0.02+ (0.01)	-0.004 (0.01)	0.01 (0.01)	103.17 (71.12)	0.01 (0.01)	92.16 (70.55)	-0.02 (0.06)	-0.03 (0.06)	-0.06*** (0.01)	-260.01 (818.33)	-0.06*** (0.01)	-222.5 (817)
Trend	0.01 (0.001)	0.01 (0.001)	0.01*** (0.001)	-7.46 (10.28)	0.005*** (0.001)	-13.73 (10.93)	0.001 (0.002)	0.002 (0.002)	-0.001* (0.0003)	-79.34*** (24.46)	-0.0005* (0.0003)	-82.83*** (25.15)
Year Dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R <sup>2</sup>	0.91	0.91	0.94	0.04	0.94	0.05	0.55	0.56	0.80	0.07	0.80	0.07

\*\*\*:  $p < 0.001$ , \*\*:  $0.001 \leq p < 0.01$ , \*:  $0.01 \leq p < 0.05$ , + indicates  $0.05 \leq p < 0.1$ .

## VI. CONCLUSION

This paper studies network effects on Kiva by testing the existence of a positive feedback loop between the number of active lenders and the amount (in dollar value or number) of open loans. We identify strong and positive network effects during Kiva's initial stage of growth. The network effects essentially disappear in the latter period, when Kiva has reached greater stability. Our results suggest that network effects are particularly important during the initial growth phase of a marketplace platform. As growth abates and competition becomes fierce, the importance of network effects declines and other tactical, behavioral and competitive factors play an increasing role. This difference between the early growth and stability periods is largely ignored in the literature.

What are the implications for the way marketplaces such as Kiva are deployed and managed? Early on, network effects are all-important, customer acquisition and speed are key success factors, and the primary objective is to grow and achieve critical mass. However, the marketplace cannot rest on its laurels following its initial growth. Rather, the network effects weaken or even disappear, forcing the marketplace to engage in constant analysis, exploration and optimization. In the particular case of Kiva, social and behavioral factors such as the ones studied in [15] are key drivers of user behavior, and the marketplace has to dynamically optimize its features so as to keep attracting new users and increase the engagement of existing users.

Our analysis is preliminary as it has a number of limitations. First, one may use other variables to study the prevalence of network effects. Second, our results are based on a single research site, and it is worth examining to what extent they extend to other network settings and marketplaces. Further, more sophisticated econometric techniques may be used to study the drivers of marketplace adoption. These extensions provide fruitful avenues for future research.

## REFERENCES

[1] H. Mendelson, "Platform business models: Text and case studies," *Electronic Business Case Collection, Kindle Edition*,

- <https://www.amazon.com/Platform-Business-Models-Electronic-Collection-ebook/dp/B078H3CDW9>, 2017.
- [2] M. L. Katz and C. Shapiro, "Systems competition and network effects," *Journal of economic perspectives*, vol. 8, no. 2, pp. 93–115, 1994. doi: 10.1257/jep.8.2.93
- [3] A. Hagiui and J. Wright, "Multi-sided platforms," *International Journal of Industrial Organization*, vol. 43, pp. 162 – 174, 2015. doi: <https://doi.org/10.1016/j.ijindorg.2015.03.003>. <http://www.sciencedirect.com/science/article/pii/S0167718715000363>
- [4] J. Rohlfs, "A theory of interdependent demand for a communications service," *The Bell Journal of Economics and Management Science*, pp. 16–37, 1974. doi: 10.2307/3003090
- [5] E. Brynjolfsson and C. F. Kemerer, "Network externalities in microcomputer software: An econometric analysis of the spreadsheet market," *Management Science*, vol. 42, no. 12, pp. 1627–1647, 1996. doi: 10.1287/mnsc.42.12.1627
- [6] T. H. Hannan and J. M. McDowell, "The determinants of technology adoption: The case of the banking firm," *The RAND Journal of Economics*, pp. 328–335, 1984. doi: 10.2307/2555441
- [7] C.-P. Lin and A. Bhattacharjee, "Elucidating individual intention to use interactive information technologies: The role of network externalities," *International Journal of Electronic Commerce*, vol. 13, no. 1, pp. 85–108, 2008. doi: 10.2753/JEC1086-4415130103
- [8] ICE Benchmark Administration Limited (IBA), "Ice bofam emerging markets corporate plus index effective yield [bamlemcbpiey]," 2019, retrieved from FRED, Federal Reserve Bank of St. Louis, <https://fred.stlouisfed.org/series/BAMLEMCBPIEY>.
- [9] G. Madden, C.-N. Grant, and B. Dalzell, "A dynamic model of mobile telephony subscription incorporating a network effect," *Telecommunications Policy*, pp. 133–144, 2004. doi: 10.1016/j.telpol.2003.12.002
- [10] B. Lloyd and M. Surana, "Online marketplaces for loans are growing rapidly. should banks be worried?" <https://www.hardingloevner.com/fundamental-thinking/online-marketplaces-for-loans-are-growing-rapidly-should-banks-be-worried/>, accessed: 2019-05-07.
- [11] Kiva Blog, "Expiring loans," 2012, retrieved in May 2019 online from <https://pages.kiva.org/blog/qa-expiring-loans-credit-limits-and-the-evolution-of-kiva>.
- [12] —, "Supply and demand," 2014, retrieved in May 2019 online from <http://blog.kiva.org/supply-and-demand#findingtrouble>.
- [13] M. Feldman, C. Papadimitriou, J. Chuang, and I. Stoica, "Free-riding and whitewashing in peer-to-peer systems," *IEEE Journal on selected areas in communications*, vol. 24, no. 5, pp. 1010–1019, 2006. doi: 10.1109/JSAC.2006.872882
- [14] H. White, "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity," *Econometrica*, vol. 48, no. 4, pp. 817–838, 1980.
- [15] H. Mendelson, K. Moon, and Y. Shen, "Behavioral and social effects in a crowdfunding marketplace," *Working Paper, Graduate School of Business, Stanford University*, 2019.

# The Approach to Applications Integration for World Data Center Interdisciplinary Scientific Investigations

Grzegorz Nowakowski

Department of Automatic Control and Information  
Technology, Faculty of Electrical and Computer  
Engineering, Cracow University of Technology  
Cracow, Poland  
gnowakowski@pk.edu.pl

Kostiantyn Yefremov

World Data Center for Geoinformatics and  
Sustainable Development, Kyiv, Ukraine  
k.yefremov@wdc.org.ua

Sergii Telenyk

Department of Theoretical Electrical Engineering  
and Computer Science, Faculty of Electrical and  
Computer Engineering, Cracow University of  
Technology Cracow, Poland  
stelenyk@pk.edu.pl

Volodymyr Khmeliuk

Department of Automation and Control in Technical  
Systems National Technical University of Ukraine  
“Igor Sikorsky Kyiv Politechnic Institute” Kyiv,  
Ukraine  
hmelyuk@gmail.com

**Abstract**—The approach to applications integration for World Data Center (WDC) interdisciplinary scientific investigations is developed in the article. The integration is based on mathematical logic and artificial intelligence. Key elements of the approach - a multilevel system architecture, formal logical system, implementation – are based on intelligent agents interaction. The formal logical system is proposed. The inference method and mechanism of solution tree recovery are elaborated. The implementation of application integration for interdisciplinary scientific research is based on a stack of modern protocols, enabling communication of business processes over the transport layer of the OSI model. Application integration is also based on coordinated models of business processes, for which an integrated set of business applications are designed and realized.

**Index Terms**—research, application integration, business processes, mathematical logic, formal logic, inference mechanisms, multi-agent systems, protocols, software agents

## I. INTRODUCTION: PARTICULARITIES OF WDC APPLICATION INTEGRATION

IN VIEW of the globalization of the economy and social life, National Science must integrate into the world and European organizations that promote the consolidation of research and consequently the development of scientific activity [2]. This is a very important process since there is an urgent need for interdisciplinary research, primarily for the assurance of sustainable development globally and regionally [3]. However, effective implementation of interdisciplinary research requires the creation of appropriate conditions for information exchange in the process of solving scientific problems. The scientific and technical progress that in its time facilitated the creation of information and communication technologies (ICT) nowadays benefits greatly from them. They are developing rapidly, covering new spheres of human activity and enhancing performance. Yet only field specialists can use ICT in a rational way, whereas the need for efficient information exchange within interdisciplinary research can only be met through rational ICT [19] application by specialists with deep knowledge in their areas of expertise [4].

The development of the information technologies (IT) domain that is experiencing qualitative changes related to ICT

Presented results of the research, which was carried out under the theme No. E-3/586/2018/DS, were funded by the subsidies on science granted by Polish Ministry of Science and Higher Education.

strengthening has created conditions for distributed computation and the efficient use of information and other resources. Consolidation of resources and the introduction of virtualization technologies are contributing to the process of substituting local solutions with distributed ones that allow the comprehensive use of all computing powers and data storage systems linked into a global network, thus granting access to accumulated information resources. The service approach formed on the basis of communication services has spread to the infrastructure, software development tools, and applications. The emergence of a wide range of new types of services, especially content-based ones, has led to the convergence of services and the formation of a generalized concept of information and communication services (ICS). The number of ICS providers has grown rapidly and convergent providers have emerged. The wide functionality, high quality, and moderate price of new services rendered by providers allow businesses to abandon the development of in-house or IT infrastructure and to use a wide variety of available ICS for component-based design of their information and telecommunication systems (ITS).

However, unified access to services is becoming a condition for the efficient use of the advantages of distributed systems and the possibilities of service-oriented technologies. At the same time, the formation of a new democratic IT environment, in which even small businesses can render services, has naturally been accompanied by the use of various tools and access technologies [5]. Therefore, historically the IT environment is heterogeneous, hence user access to its resources is to some extent complicated or at least inconvenient.

The same situation is characteristic of scientific activity in particular. For example, the World Data Centers (WDC) system created in 1956 under the aegis of the International Council for Science (ICSU) ensures collection, storage, circulation, and analysis of data obtained in various science areas [2]. During its existence, the WDC system has accumulated a lot of data and applications that may be used to solve the challenging problems of social development. They are one of the most powerful information resources used by hundreds of thousands of scientists, and the demand for it is increasing in proportion to the need for interdisciplinary research related to sustainable development, the solution of

urgent environmental protection issues, etc. However, problems related to the incompatibility of legacy applications caused by architecture differences, the variety of data presentation formats, and other factors prevent the effective use of such WDC resources.

Promising architectural solutions are being developed and gradually implemented in the ICT field, such as Next Generation Network (NGN) and Next Generation Service Overlay Network (NGSON) [6], that ensure the interaction of various transport layer technologies. Yet there is a need for a comprehensive integration of resources from various sources, and ICT developers' efforts should seek to enable scientists working in various domains to use the accumulated resources based on their areas of expertise, and not on the IT particularities. The sources, comprehensive access to which it is reasonable to ensure, include databases, websites and portals, various legacy file management systems, and data repositories structured according to various models.

Nowadays there exist more than 50 WDCs that for more than 50 years have created a system for data accumulation, analysis, processing, and international exchange. WDCs' powerful data storage systems retain huge volumes of astronomical, geophysical and other scientific data. WDCs' servers process this data using numerous and various applications powered by different technologies.

Certainly, from the point of view of scientists working on various resource-intensive problems, it is important to have access not to a large disordered system of possibilities, but to an integral complex of data and application sources intended to meet their specific needs, with a user-friendly interface that does not require any special IT knowledge. Nevertheless, the system for WDC data accumulation, analysis, processing, and international exchange does not provide such access. Furthermore, it was not designed for the growing level of scientific society's requirements and is not versatile enough to be used in interdisciplinary research. Therefore, a new interdisciplinary structure was created in 2008 — the World Data System (WDS) — to develop and implement a new coordinated global approach to scientific data, which guarantees omni-purpose equal access to quality data for research, education, and decision making. The new structure will have to solve the accumulated tasks, primarily the unification of formats and data transfer protocols, assurance of convenient access to data, and the organization of scientific data quality control [2].

Integrating formerly independent systems for the accumulation, storage, and processing of WDC data on a new advanced integration basis will allow a considerable enhancement of their overall efficiency. The creation of such a system will provide scientists with convenient centralized access to formerly separate resources, facilitating and quickening scientific and research activity around the world.

## II. RELATED WORK

An overview of existing integration solutions in all the mentioned aspects, starting with the integration technology aspect has been done. The existing solutions are powered by technologies for the creation, functioning, and development of distributed systems. Notwithstanding the differences of various technologies for the creation of distributed service-oriented systems, the overall principles and theoretical and methodological approaches are always similar. Independent services should be registered, described, and provided with the possibility to communicate transparently with clients and

with each other. Furthermore, networking interaction requires us to determine protocols for all levels of the OSI model. To do so, the corporate ITS standard structure with services intellectualization operations may be used, as suggested in [8], taking into account international, state, and branch standards, and corporate documents, first of all [9]. The structure consists of two parts, where the first covers the traditional four levels of standards of the TCP/IP protocols stack, and the second covers the user applications in accordance with the ITS class, destination, and services intellectualization operations. Figure 1 presents an example of the corporate standard structure.

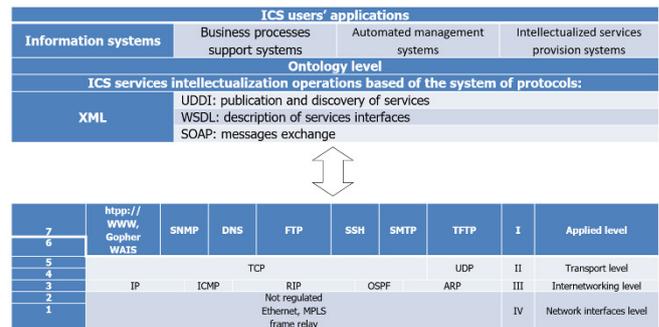


Fig. 1 Stack of services interaction protocols

For convenience, the figure presents the levels of the international standard for interaction of open OSI systems and their correlation with the corresponding levels of the TCP/IP protocols stack. The ITS classification used is proposed in [8]. It allows the systematization of various ICS by the level of users' requirements and queries, practical needs, and professional training, with no limitations imposed on the ITSs' functionality, attributes, or operations. The changes introduced apply to the first part of the structure and to the detailing of protocols that power the intellectualized interaction of applications in the course of solving users' more complex problems.

The services registry is maintained by the Universal Description, Discovery and Integration (UDDI) technology [10] that allows both people and client-programs to publish information on services and search for a required service.

The Web Service Description Language (WSDL) that, according to the W3C definition, constitutes an XML format for the description of networking services as a set of operations working with document- or procedure-oriented information through messages [11], is used for the unified description of services, which allows them to be used independently from the programming language. WSDL documents, by virtually creating a unified layer that allows the use of services created on the basis of various platforms, describes the service interface, URL, communication mechanisms 'understood' by the service, methods provided by it with corresponding parameters (type, name, location of the service Listener), and the service messages structure.

To implement the key part of the interaction—messages exchange—one of the widespread technologies may be used, and the selection should be determined by compliance with the requirements of the system for data accumulation, processing, and exchange:

**SOAP:** a user-friendly technology that is easy to use with the Business Process Execution Language (BPEL), which ensures interaction of distributed systems irrespective of the object model, operating system, or programming language. Data is transferred as special-format XML documents [12];

**CORBA:** a mechanism created to support the development and deployment of complex object-oriented applied systems, which is used for the integration of isolated systems, allowing programs that are developed in different programming languages and working in different network nodes to interact with the same ease as if they were in the address space of a single process. Such interaction is ensured through unified construction of their interfaces using a special-purpose declarative Interface Definition Language (IDL). At the same time, the interface and the description thereof do not depend on the operating systems or the processor architecture either [13];

**REST** (Representational State Transfer) is a style of software architecture for distributed systems, which is used to create web services. A global URL identifier unambiguously identifies each unit of information in an unvarying format. Data is transferred without any layers [14];

**WCF** (Windows Communication Foundation) is Microsoft's platform designed to create applications that exchange data through the network and are independent from style and protocol [15].

Although nowadays the most acceptable choice seems to be SOAP, based on which the interaction presented in Fig. 1 is performed, all the technologies allow the integration of services of various origin [16] – [17].

To take into consideration service-based, process-functional, and component-based approaches to the design and maintenance of users' solutions, the following well-known architectural means of organizing shared functionality are used:

Web Service Choreography is the approach that determines the protocols of web services' interaction to perform a single global task by performing parts thereof. The role assumed by the service determines its model of exchanging messages with other services. This method shows high efficiency for small tasks, but with increased complexity of tasks the number of services involved grows rapidly, solutions become too massive, and the efficiency drops quickly [18].

Based on the abovementioned technologies, a few solutions have been worked out, which may be used to create applied systems for the accumulation, processing, and exchange of scientific data in different areas. The best known among them are the ESIMO and GEOSS systems. They are quite widespread, although they have a number of drawbacks in terms of processing heterogeneous information.

One of the most crucial issues is that of developing mathematical models, methods, and means for the integration of various applications, both legacy ones based on traditional technologies and client-server and web-oriented technologies. The article proposes an approach to applications integration using the mathematical logic instrument and artificial intelligence theory for interdisciplinary research through the example of the WDC system functioning. The specificity of the applications integration for interdisciplinary research is

that the coordination of business processes models is not required because it is, in fact, substituted with schemes for performing users' tasks. In general, the applications integration is performed on the basis of coordinated business processes models, and the integrated complex of business applications is intended to support them.

### III. DEFINITION OF THE APPLICATION INTEGRATION PROBLEM FOR INTERDISCIPLINAR REASARCH

Mathematical models and methods as the basis for a holistic solution should be created to power data storage and processing centres, which will provide users with versatile possibilities at the level of information integration and servers availability. The mentioned models and methods should be devoid of the drawbacks characteristic of the algorithmic approach, related to the need to reprogram data processing algorithms upon the emergence of new data types and changes in the implemented algorithms or the emergence of new ones. At the data processing level, such a solution should ensure:

- computations distribution and the use of remote hardware resources;
- versatility and adaptation to the system load;
- system usage space;
- data supply by remote client or service;
- availability of intellectual data processing means directly to the end user with no special knowledge or skills;
- fast and simple integration of data and applications of various global information systems

### IV. THE APPLICATION INTEGRATION SYSTEM ARCHITECTURE

The principal tenet for the creation of a distributed system is the method of organizing services interaction. Of the two known main ways of services interaction - orchestration and choreography - the more efficient for the WDC systems is the former. Indeed, orchestration that aggregates basic services into hierarchically integrated systems, subordinating them to administrators – 'orchestrator' services - allows services to be unified by attributes that are convenient to WDC (science area, functionality, regional location, etc.), and to provide orchestrator services with powers in accordance with the international data exchange policies. Thus, it is possible to line up a simple and effective system, in which the search for the necessary services or the construction of their composition for the complex queries inherent in interdisciplinary research will not require a lot of time, as every Orchestrator Service has information on the functionality of inferior services, and in addition they can coordinate their possibilities in the process of the planning and execution of users' queries. Plugging new services into the system will not present a particular problem either, as doing so will only require the service description to be laid down, entered into the registry, and assigned to a certain orchestrator service.

Such an approach will work when users know exactly their information needs and have corresponding knowledge and skills to compile chains of queries to the known orchestrators. What is more important is that the approach will provide the opportunity to work with the system for users who cannot initiate services, by determining the sequence of their work and specifying execution and interaction parameters. Users only have to know how to formulate their needs in terms of a particular subject domain. In this case, the association of services and the organization of their cooperation require an intellectual constituent. For this purpose, intellectual agents are used, which constitute the system core, implementing its functioning logic, which promotes the formation of queries, plans their execution, and organizes basic services interaction. The introduction of intellectual agents as orchestrator services into the hierarchical structure forms a two-tier system; the bottom level consists of services performing basic tasks, the top level consists of intellectual agents that orchestrate basic services. By using the registries of subordinated basic services and their functionality and by interacting with each other, interconnected intellectual agents implement methods of logical inference. The inference result is the composition of basic agents' operations that allow user-defined tasks to be performed. This operations composition, or proof, is transferred to the lower level, where the operations necessary to solve the user's task are performed. At this point, control is handed over to the lower level agents that only return the final result to be sent to the user.

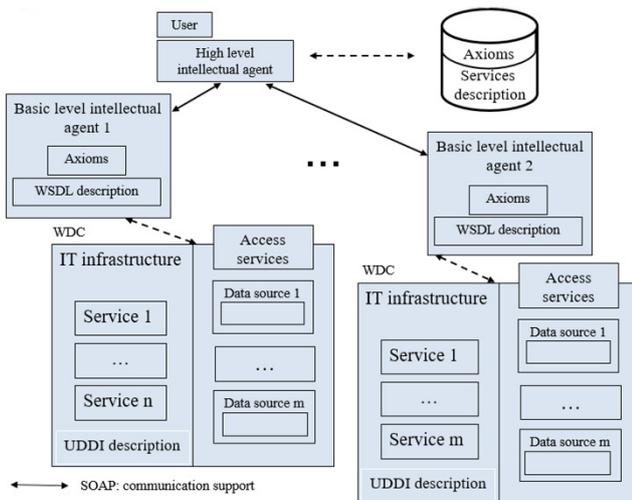


Fig. 2 System components interaction scheme

The user's query itself is performed upon formation of the proof containing references to one or several pointers to data sources and IDs of methods that the system services should apply to the data. Some queries do not require pre-processing of data, since the execution of services methods is sufficient upon receipt of information from the data source. Others require a certain sequence of prior data operations, and as a result, the execution of other services methods. The necessary connections are described in the axioms entered into the knowledge base upon registration of respective services in the system. These operations yield the final product of the system - data that are the solution to the user's problem. Since the system operation is determined by the user's query, the

system interface should help them make a query in familiar terms.

The proposed solution uses UDDI to create the service registries and WSDL for the unified description thereof. The key part—messaging—is implemented using SOAP, and services orchestration—using BPEL. Agents are described by means of JADE. The system components interaction scheme is presented in Fig. 2.

The description of the interface (WSDL) of every low-level service registered in the system and of related axioms is input into the system knowledge base and directly into the knowledge base of the Controller Agent to which the new service will be subordinated. The Controller Agent is selected in accordance with the policies adopted in the WDC system, for example, by the criteria of the geographical location of the deployed agent, in order to minimize data exchange.

At the agent level, the orchestrator agent (or several), having received the user's query, interrogates the agents in order to search for the necessary services and available resources. The user's query is processed taking into account the services of the system, their functionality (described by the axioms of the system), and the inference rules defined by logical formalism. The resulting proof is transferred to the lower level to be implemented.

At the services level, each agent invokes the required services from its set and sends processed data according to the action chains. This process is carried out in accordance with the task solution tree reconstructed by the solution tree reconstruction mechanism based on the proof.

To realize this interaction of services in the WDC system, the most appropriate is the logical approach, which allows both the level of the abovementioned requirements to be reached and the drawbacks of a traditional algorithmic approach to be eliminated. Indeed, it is only required to develop the inference method and the solution tree reconstruction mechanism that will implement the query formation processes, plan their execution, reconstruct and implement the task solution scheme. The logical approach is the most appropriate one to create and describe these constituents. The logical approach implementation requires one to:

- describe the existing applications and their functional capacities in the formal language;
- formulate the inference rules;
- determine the inference method;
- develop an algorithm of the user's query execution tree reconstruction based on the proof;
- implement these methods and mechanisms in the agents of the system.

## V. THE FORMAL LOGICAL SYSTEM

Let us describe the formalism upon which the program system that will ensure solution of the formulated problem will be built. We will take the first order clausal logic for a base and describe the formal system language in accordance with the structural elements determined in [3].

**Symbols: service:** ( ), [ ], { }, :, <, >, ;

**constant:**

- 1) *individual, of primary types* (int, real, char, bool) -  $a_1^1, a_2^1, \dots, a_1^2, a_2^2, \dots$  where each constant  $a_i^k$  pertains to type (primary type)  $k$ ; *structural type* (construct) -  $c_1, c_2, \dots$ ; *procedural type* (method) -  $d_1, d_2, \dots$ ; *objective type* (problem, entity, relation) —  $e_1, e_2, \dots$ ;
- 2) *functional i-place*, for individuals of type  $k$  -  $h_1^1, h_2^1, \dots, h_1^2, h_2^2, \dots$ ;
- 3) *predicate i-place*, for individuals of type  $k$  -  $A_1^1, A_2^1, \dots, A_1^2, A_2^2, \dots$  (this class includes taxonomic, relational and other predicates, as well as traditional relations, at least equality = and order  $\geq$ );

**variable:** for individuals of type  $k$  -  $x_1^1, x_2^1, \dots, x_1^2, x_2^2, \dots$ ,

where every variable  $x_i^k$  pertains to type  $k$ ;

**logical:**  $\neg, \wedge, \vee, \leftarrow, \exists, \forall, \leftrightarrow$

**Individual terms of type  $k$ :**

- 1) each individual constant  $a_i^k$  of type  $k$  is an individual term of type  $k$ ;
- 2) each free variable  $x_i^k$  for individuals of type  $k$  is an individual term of type  $k$ ;
- 3) if  $h_i^j$  is a certain functional constant for individuals of type  $k$  and  $\tau_1, \dots, \tau_j$  are terms for individuals of type  $k$ , then  $h_i^j(\tau_1, \dots, \tau_j)$  is an individual term of type  $k$ ;
- 4) there are no other individual terms of type  $k$ .

The terms obtained by applying construction rules 1 or 2 of the definition will be called primary, and all the others—complex.

**Formulas for individuals:**

- 1) if  $A_i^j$  is a predicate constant for individuals and  $\tau_1, \dots, \tau_j$  are terms for them, then  $A_i^j(\tau_1, \dots, \tau_j)$  is the atomic formula for individuals;
- 2) the atomic formula for individuals is the formula for them;
- 3) there are no other formulas for individuals.

Hereinafter we consider that the system contains an omni-purpose transformer of formulas into the traditional for the clausal form (we are talking about the Horn clauses) view with a single  $\rightarrow$  symbol, atomic formulas to its left and right, and an implicit quantifier  $\forall$ .

Specifiers, preconditions, post-conditions, specifiers of methods, specifiers of problems, clause, the system's knowledge and inference rules are presented in detail in [1].

## VI. THE INFERENCE METHOD

The method proposed in [2] based on analogy and types of assertions has been used. A detailed description of this method can be found in [1], [4], [7].

The inference mechanism work algorithm is presented in Fig. 3.

To improve the inference mechanism three known elements were integrated:

- 1) a multiset of literals (atomic formulas) (which will be called a multi clause (m-clause));
- 2) the ordered linear;
- 3) typification abstraction (to manage an ordered linear proof).

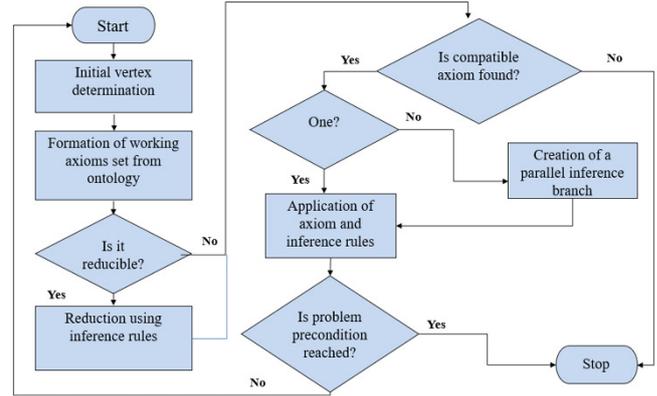


Fig. 3 Inference mechanism work algorithm

## VII. THE MECHANISM OF SOLUTION TREE RECONSTRUCTION

To obtain from the proof generated by the inference mechanism a functional sequence of actions to be used by our system, taking account of services features and nature, the solution scheme reconstruction mechanism should be initiated [1].

Condition: proof Result = $\langle V, T \rangle$ . To find: $G = \langle V, E, \Theta \rangle$	
Step	
1	Identification of the terminal triple—the triple that contains vertex $k$ that is neither the first nor the second component of any of the triples of set $T$ .
2	In the selected triple, the initial vertex is the post-condition vertex, the second is the method vertex, and the terminal is the precondition vertex. The method that turns the precondition into the post-condition is determined by the unambiguous correspondence to the vertex formulas of the post-condition and precondition of the method axiom. The vertices are connected with the edges: precondition to method and method to post-condition. If the precondition of the method axiom consists of a combination or intersection of several elements, a data vertex is inserted between the precondition and the method vertices, which groups the preconditions according to the relevant inference rule. If the terminal triple only contains two vertices, it means that the inference rule was used to convert the vertex clauses, and the place of the method vertex in this tree branch is occupied by the transformation data vertex that corresponds to the inference rule used, with the corresponding post-condition vertices.
3	The vertices, edges, and their correspondence are introduced into $G$ .
4	Having extracted the processed triple from the set of triple vertices $T$ , we repeat step 2. If the set is empty, the tree reconstruction is completed.

Fig. 4 Algorithm for vertex scheme reconstruction

The solution scheme constitutes a connected directed graph with no oriented cycles with parallel directed paths from the root to the vertices; it has three types of vertices, and is specified by triple  $G = \langle V, E, \Theta \rangle$ , where  $V = V_1 \cup V_2 \cup V_3$ ,  $V_1$  is a set of method vertices,  $V_2$  is a

set of precondition vertices and post-condition vertices,  $V_3$  is a set of data vertices in which data is merged or split;  $E$  is a set of edges;  $\Theta$  is a subset of Cartesian product  $E \times V \times V$ , which determines the correspondence of edges to pairs of vertices. The scheme determines the sequence and the correspondence according to the data of actions to be performed by the data processing system's executive mechanisms to obtain the result desired by the user [1].

Algorithm for vertex scheme reconstruction was presented on Fig. 4.

The reconstructed solution tree is fed to the actuator input. The implementation of a particular method specified in the solution tree is represented by a construct that describes the input data (entities, connections, relations between them), method preconditions, method post-conditions, and output data. The solution tree branches downstream of the data-splitting vertex can be executed in parallel until they reach the data-merging vertex, where, after completing all the parallel branches involved in merging, they continue to be executed consecutively. The work algorithm of the solution tree reconstruction mechanism is presented in Fig. 5.

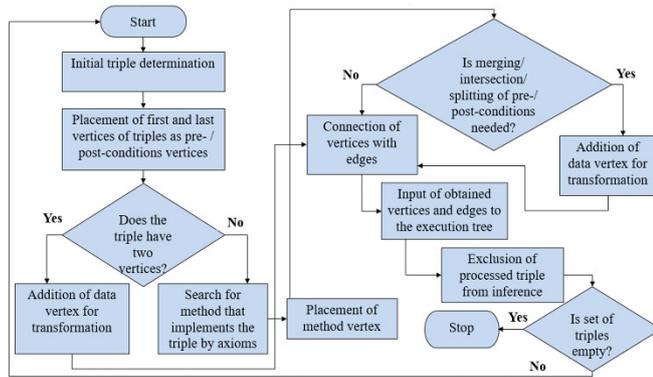


Fig. 5 Work algorithm of solution tree reconstruction mechanism

## VIII. APPLICATION OF LOGICAL APPROACH TO PROBLEM SOLUTION

The approach efficiency on the real scientific task of calculating the component of life safety and indicating critical values of threat indicators for analyzing the sustainable development of the regions of Ukraine has been demonstrated.

The life safety component formula:  $Csl = \sqrt[3]{\sum_0^n Threat^3}$ , where the threat value is normalized by formulas (2) or (3). The purpose of indicating critical values of threat indicators is to determine the priority in consideration thereof in the decision-making process on the level of a single region and the entire country in order to mitigate the impact of threats on sustainable development.

Suppose that for every administrative unit  $i = \overline{1, n}$  there is a set of values  $\langle x_{i,1}, x_{i,2}, \dots, x_{i,m} \rangle$  of indicators  $X_j, j = \overline{1, m}$ , which characterize the negative impact of certain phenomena on the sustainable development processes in the economic, social, and ecological spheres. Such indicators, whose essence and

composition are determined by experts, will be called threat indicators.

Given the content of the critical values indication problem, the following characteristic function must be determined:

$$\Psi(x_{i,j}) = \begin{cases} 0, & \text{if value of } x \text{ is not critical} \\ 1, & \text{if otherwise} \end{cases} \quad (1)$$

where  $i = \overline{1, n}, j = \overline{1, m}$ .

It is clear that the determination of function  $\Psi(x_{i,j})$  must be based upon certain criteria that take into consideration exceeding by  $x_{i,j}$  a hazardous limit, the relative position of region  $i$  in the indicators rating  $X_j$  compiled for the comparison groups and for the entire country, and the degree of "hazard" of value  $x_{i,j}$  in comparison to values of other indicators for region  $i$ .

To account for the relative position of the region in the entire country's ratings, the following criterion is used:

$$R_{i,j} = \left( 1 + e^{\frac{a - x_{i,j}}{b}} \right)^{-1} \quad (2)$$

if higher values of indicator  $X_j$  correspond to a higher impact of the respective threat on the sustainable development, and:

$$R_{i,j} = 1 - \left( 1 + e^{\frac{a - x_{i,j}}{b}} \right)^{-1} \quad (3)$$

if lower values of indicator  $X_j$  correspond to a higher impact. In formulas (2)-(3), parameters  $a$  and  $b$  are calculated by the following formulas:

$$a = \overline{X_j} = \frac{1}{n} \sum_{i=1}^n x_{i,j}, \quad b = \sigma(X_j) = \sqrt{\frac{\sum_{i=1}^n (x_{i,j} - \overline{X_j})^2}{n}} \quad (4)$$

Criterion  $R_{i,j}$  is a dimensionless number that assumes values within  $[0, 1]$ . Values of around 0.5 correspond to average values of  $X_j$  in the selection, and values higher than 0.75 correspond to values that exceed the average ones by more than a standard deviation. In this case, the characteristic function (1) taking account of one criterion  $R_{i,j}$  may be expressed as follows:

$$\Psi_R(x_{i,j}) = \begin{cases} 0, & R_{i,j} < 0,75; \\ 1, & R_{i,j} \geq 0,75. \end{cases}$$

Criterion  $P_{i,j}$  that takes into account a region's relative position in the comparison group may be calculated by formulas (2)-(3) taking account of the fact that parameters  $a$

and  $b$  are calculated by formula (4) independently for each comparison group.

Criteria  $R_{i,j}$  and  $P_{i,j}$  are dimensionless numbers and are of similar nature and can therefore be aggregated through a weighted sum:

$K_{i,j} = w_R R_{i,j} + w_P P_{i,j}$ ;  $w_R + w_P = 1$ , where weighting factors  $w_R$  and  $w_P$  are determined by experts.

Thus, for region  $i = \overline{1, n}$  we have a set  $\langle K_{i,1}, K_{i,2}, \dots, K_{i,m} \rangle$  of values of the aggregated criterion that accounts for a region's relative position in ratings compiled for comparison groups and for the entire country. Now, among values  $K_{i,j}$ ,  $j = \overline{1, m}$ , the worst must be determined, which may also be performed by formula (2):

$$I_{i,j} = \left( 1 + e^{\frac{a - K_{i,j}}{b}} \right)^{-1}$$

where parameters  $a$  and  $b$  are calculated in selection  $K_{i,j}$ ,  $j = \overline{1, m}$ .

For values of criterion  $I_{i,j}$  the same remarks apply as for criterion  $R_{i,j}$ . Therefore, characteristic function (1) may be expressed as follows:

$$\Psi_I(x_{i,j}) = \begin{cases} 0, & I_{i,j} < 0,75; \\ 1, & I_{i,j} \geq 0,75. \end{cases}$$

Thus, values  $\Psi_I(x_{i,j}) = 1$  correspond to the highest priority of attention to be paid to the value of indicator  $X_j$  in the administrative decision-making process on the level of single region  $i$ .

## IX. CONCLUSIONS

Based on the analysis of existing data centres, their equipment and software, a high quality solution is offered that provides a simple and flexible way to integrate heterogeneous information systems and their services into the World Data System. One of the key features of the proposed solution is the automation of the algorithm construction of the actions sequence that executes users' queries. A logical formalism has been created to describe this solution, and on its basis an inference method and a solution tree reconstruction mechanism have been developed.

The analysis of available technologies used for the implementation of distributed systems allowed the use of such a set of software solutions for practical implementation of this solution: UDDI for creating a registry of services entered into the system; WSDL for a unified description of services; SOAP for exchanging notifications between services; BPEL for the overall coordination of services. Intellectual agents can be implemented using JADE.

The implementation of the proposed solution will provide an opportunity to use all the integrated computing capacities and data storage systems of the World Data System in a comprehensive manner. Thus, users will be able to easily gain access to all the necessary resources and services available to the system.

## REFERENCES

- [1] S. Telenyk, G. Nowakowski, K. Yefremov and V. Khmeliuk, "Logics based application integration for interdisciplinary scientific investigations", 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), pp. 1026-1031, Bucharest, 2017. DOI: 10.1109/IDAACS.2017.8095241
- [2] M. Z. Zgurovsky, A. D. Gvishiani, K. V. Yefremov and A. M. Pasichny, "Integration of the Ukrainian science into the world data system", Cybernetics and Systems Analysis: Volume 46, Issue 2 (2010), pp. 211-219. DOI: 10.1007/s10559-010-9199-9
- [3] M. Z. Zgurovsky, A. O. Boldak, K. V. Yefremov and others, "Analysis of Sustainable Development – Global and Regional Contexts", International Council for Science (ICSU) and others. – K.: NTUU «KPI». – Part 2. Ukraine in Sustainable Development Indicators (2011-2012). – 232 p, 2012.
- [4] O. Pavlov, S. Telenyk. *Algorithmization and IT in management*, Kyiv: Technics, 2002 – 320 p.
- [5] Data Integration Information, quick view on world of data, (online) homepage at: <https://www.dataintegration.info/>
- [6] M. Ulema et al., "Next generation service overlay networks (NGSON)", IEEE Communications Magazine 50(1):52-53, 2012, DOI: 10.1109/MCOM.2012.6122532.
- [7] A.Y. Levy, "Logic-Based Techniques in Data Integration", In: Logic Based Artificial Intelligence, Edited by J. Minker. Kluwer Publishers, 2000.
- [8] P. P. Maslianko, "Fundamentals of the methodology of system design of information and communication systems", Naukovi Visti NTUU "KPI," No. 6, 54–60 (2007).
- [9] OMG Systems Modeling Language, (online) homepage at: <https://www.omg.org/spec/SysML/About-SysML/>
- [10] UDDI Version 3.0.2 Specification, (online) homepage at: [http://uddi.org/pubs/uddi\\_v3.htm](http://uddi.org/pubs/uddi_v3.htm)
- [11] J. Greer, "Web Services Description Language: 55 Most Asked Questions: What You Need to Know", Emereo Publishing, 2014
- [12] SOAP Version 1.2 Part 1: Messaging Framework (Second Edition), (online) homepage at: <https://www.w3.org/TR/soap12-part1/>
- [13] About the common object request broker architecture specification version 3.3, (online) homepage: <https://www.omg.org/spec/CORBA>
- [14] G. Nowakowski, "Rest Api safety assurance by means of HMAC mechanism", Information Systems in Management, Vol. 5, No. 3, pp. 358-369, 2016.
- [15] Windows Communication Foundation Architecture Overview, (online) homepage at: <http://msdn.microsoft.com/en-us/library/aa480210.aspx>
- [16] S. Graham, et al., "Building Web Services with Java: Making Sense of XML, SOAP, WSDL, and UDDI (2nd Edition)", Sams Publishing; 2 edition, 2004
- [17] S. El-Seoud, H. El-Sofany, M. Abdelfattah, M. Reham, "Big Data and Cloud Computing: Trends and Challenges", International Journal of Interactive Mobile Technologies, 2017, Vol. 11 Issue 2, p34-52, 19p, 3 Diagrams; DOI: 10.3991/ijim.v11i2.6561
- [18] J. Laznik, Y. Mannari, R. Dhruv, BPEL and Java Cookbook: Over 100 Recipes to Help You Enhance Your SOA Composite Applications with Java and BPEL, Birmingham, 2013
- [19] E. Ziemba, "The ICT Adoption in Government Units in the Context of the Sustainable Information Society", 2018 Federated Conference on Computer Science and Information Systems, pp.725–733. DOI: 10.15439/2018F116



# Information Systems Development and Usage with Consideration of Privacy and Cyber Security Aspects

Janusz Jabłoński  
Uniwersytet Zielonogórski  
ul. prof. Z. Szafrana 4a  
65-516 Zielona Góra, Poland  
Email: j.jablonski@wmie.uz.zgora.pl

Silva Robak  
Uniwersytet Zielonogórski  
ul. prof. Z. Szafrana 4a  
65-516 Zielona Góra, Poland  
Email: s.robak@wmie.uz.zgora.pl

**Abstract**—One of the contemporary problems, and at the same time a challenge, with development and usage of supply chain Information Systems are the issues associated with privacy and cyber security, which emerged due to new requirements of legal regulations and directives. The human factor belongs to the biggest risks within these issues. Leak of information, phishing, unauthorized access are the main problems. Also vulnerability of the systems due to new information technologies is an important topic. In this paper we discuss development and usage of Information Systems with regard to the security aspects associated to the software development lifecycle. We present our approach on examples of a user authentication process in logistics.

## I. INTRODUCTION

THE information security and cyber security are strongly associated with the technological infrastructure of computer networks and computer systems processing information. A computer system is secure if the user can rely on its functionality and the installed application software is working consequently the specifications. Developing software applications with compliance to the user requirements is not sufficient, because the developed systems should additionally be secure and consistent with the current state of law regulations.

In this paper we will approach a problem of the exchange of a vast amounts of data in supply chains with respect to the data privacy and security issues. There is the European Union's General Data Protective Directive GDPR concerning the protection of natural person with regard to the processing of personal data [1]. On the background of this regulation and of the Payment Services Directive 2 PSD2 [2], which are concerning transaction systems on financial markets, and Fintech [3], the development of the secure software business applications turns out into a great challenge. In our paper we will suggest some improvements to the IS's development process, which result from the above stated system requirements and the further implications regarding privacy and data security aspects.

The supply chain defines the network that comprehends all the organizations and activities associated with the flow

and transformation of goods from the raw material stage, through to the end user, as well as the associated information flow [4]. In our paper we will concentrate on threats and possible solutions demanded for the secure supply chain activities and flow of information.

In the inter-organizational information systems, which link the companies to their suppliers, distributors and customers, a movement of information through electronic links takes place across organizational boundaries, between separately owned organizations. It requires not only the electronic linkage in form of basic electronic data interchange systems (as for purchase orders), but also the interactions in complex cash applications and information systems or an access to shared technical databases. Thus, the problems associated with the privacy and security are also very viable in supply chains contexts.

The credibility of information as also especially the trustworthiness of the participants in supply chains is required. In transportation and logistics, in order to eliminate a possibility of the documents frauds, non-existent suppliers or recipients, an essential element of the risk elimination in supply chains is the credibility ensured by an authentication.

We believe, that the enterprise information systems being a part in a logistic supply chains should be secured during all stages of their life-cycle, and will give some guidelines for development-time and run-time of the IS.

The security concerns become additionally significant with the regulations like General Data Protection Regulation (GDPR) in the European Union and other regulations to be expected coming soon. The problem is that they use terms like "reasonable security procedures" or "appropriate practices" and do not advice what type of technology is needed to protect the personal and enterprise data. They only state generally about the responsibility of the organizations to keep data secure [5].

Therefore, as stated previously, in our paper we will analyze how to integrate the needed privacy and (cyber) security aspects into the life cycle of Information systems to

ensure the above mentioned secure procedures and appropriate practices. For this aim the rest of the paper is organized as follows.

In Section 2 we will characterize the main threats and vulnerability aspects considering the information systems security due to the usage of new emerging IT solutions and the influence of EU regulations associated with privacy and security concerns. In Section 3 we will review some aspects of cyber security and then propose some solutions applicable for developing and using information systems. The aspects of information security due to the problems with user authentication and data access control are the main topics in Section 4. In Section 5 we give some examples for conducting user authentication and show how they can support an achievement of the required privacy needs for IS of enterprises according to the EU regulations. In the last Section we conclude our work.

## II. VULNERABILITY OF INFORMATION SYSTEMS SECURITY DUE TO NEW TECHNOLOGIES

The numerous cyber attacks associated with, i.e. a stealing of the identity, the leaks of vulnerable data, or the frauds in billion of dollars yearly raised new approaches in the risk taxing, as for instance shown by Global Economic Crime and Fraud Survey for 2018 in [6]. The rethinking of the ways and approaches for development and usage of software systems and considering the proper handling of modern technologies, and also the computer networks security problems are needed. In addition, the raising numbers of mobile technology users of Smartphones and tablets with the integrated Wi-Fi equipment, and the widening popularity of operation systems like Android and iOS, caused that the mobile systems are replacing gradually the traditional computer systems.

In [7] there is a diagram depicting the percentage of the mobile OS used on market based on the report showing the growing dominance of the mobile operating systems in the last two years. The mobile devices are currently used for the e-mail checking, news viewing, the communication in social networks, and also for the payments. Operating in such environments often requires a usage and sending of vulnerable private data, such as private contact data or/and the bank account information directly with the mobile devices. It could happen that a user do not have the sufficient consciousness and knowledge of the threats caused by the neglecting of the security features on the stage of software application development.

Considering the human factor, there could be also the security risks connected with the intended conscious handling of some enterprises or programmers developing software applications, which are acting in contradiction with the users aims and also the laws regulations with the aim of processing and stealing of vulnerable user data. Examples of such behavior are known, and widely described in the Internet and include the deeds such as notorious hacker

groups, the hybrid warfare [8, 9, 10], up to the cyber troops [10, 11].

We believe that activities aimed at eliminating of vulnerabilities related with the human factor could be the particularly the forecasting of the attacks, such as attack vector [12] (i.e. email attachments, pop-up windows, deception, chat rooms, viruses and instant messages) and also regarding them possibly early already on the stage of developing software and the by usage and maintenance of IS. Such actions will be needed not only for new systems (developed from scratch), but also in maintaining already existing relative new modern and legacy systems.

Moreover, considering the vulnerabilities enumerated by the Open Web Application Security Project OWASP Foundation [13], the counteractions, or possibly elimination of the some threats is also strongly desirable by defining constrains for systems, which are using new technologies like cloud computing [14] and/or blockchain [15].

By developing software systems there will be some additional basic considerations viable for aims of their future security. To begin with gathering of user requirements and enhancing them with the law regulations related to the user privacy and data security aspects. A proper choice of the software architecture of a software system, which will be supporting the required security needs, and also the secure procedures for user authentication and system access control are recommended. There are also some additional aspects in the development phase, as for instance the usage of libraries, which are resistant to the buffer overflow, eliminating of the redundancy by avoiding linkage to the external resources, and also eliminating redundancy in computer network communication between the hosts, etc. We will consider these requirements in the following Sections.

## III. CYBER SECURITY SOLUTIONS

While the ERP systems with the embedded automated financial settlements are the constituents of supply chains, their authentication process should accomplished at a proper security level. For this reason to stay in accordance with formal requirements, the systems in logistics and supply chains should meet the requirements of the dynamic authentication in order to eliminate a possibility of hijacking or the replay attack as in case of a spoofing attack.

The authentication and access control to digital resources are the crucial elements for ensuring security in a cyberspace. The EU Regulation 2015/1502 from 8<sup>th</sup> of September 2015 defines the minimal technical specifications and the procedures for the assurance levels of electronic identification and trust services for electronic transactions at the internal Europe Union market. The regulation defines tree assurance levels [16] as:

- Low,
- Substantial, and
- High.

They should be applied for electronic identification means issued under an electronic identification scheme. Additionally in the regulation the “dynamic authentication” is defined with the meaning of “an electronic process using cryptography or other techniques to provide a means of creating on demand an electronic proof that the subject is in control or in possession of the identification data and which changes with each authentication between the subject and the system verifying the subject’s identity”.

According to this regulation for the substantial and high assurance levels in authentication mechanism, the sending of person identification data should be preceded by a reliable verification method by the electronic identification means, and its validity assured by a dynamic authentication.

For this electronic proof it is also required to be modified (to alter) with each new user authentication, as well as to be resistant to the attempts of off-line analysis.

For meeting of the above requirements included in the EU regulation, concerning the dynamic authentication mechanisms, we believe that a promising solution may be the usage of cryptography with one-time passwords or one-time key, referred to as OTP [12]. A proof for semantic security of crypto-systems constructed with regard to the rule for one-time key in cryptography was given in year 1949 by C. Shannon in [17]. Nevertheless, the research efforts regarding secure cryptographic systems implementing the OTP rules are still ongoing.

We should emphasize the fact that the systems, which implement the OTP rule are potentially resistant to the cryptanalysis with the quantum computers. Cryptography considered as resistant to the attacks by usage of quantum computing is referred as post-quantum cryptography [18].

The transaction security is also required in Fintech services, such as e-banking, e-health, etc., where the keeping anonymity and also the user authentication should be in accordance with the high assurance level. At the same time the dynamic authentication on the middle and high assurance levels, as for data which is secret and with all right reserved, is defined by the Payment Service Directive (EU) PSD2.

It is to emphasize that the above mentioned GDPR and PSD2 have been indeed introduced as regulations, but until now, the applications implementing the dynamic authentication systems conforming to these requirements are lacking.

There are some known approaches to a deal with the above problem of a secure authentication, like a research on the one-time keys in user authentication method RUBLON [19]. In the RUBLON system based on a solution given in the patent [20] is applied, and it conforms to the OTP and semantic security requirements. What is more, on the base of the solution included in the patent [20] the enterprise DCD has applied this solution in the project CryptONE (unconditional secure crypto-processor [21]), where the decryption takes place with one-time passwords.

The concept of Industry 4.0 [22] has shown some digital trends, such as the process automation and usage of the artificial intelligence in the decision processes. Therefore, in the future the authentication not only of the persons and entities, but also of the devices and the processes will be needed.

Regarding the trust, privacy and security aspects in the life-cycle of information systems with respect to the threats and vulnerabilities discussed in Section 2, below we summarize and suggest some guidelines for development and system usage (run-time) in the IS lifecycle.

In the development and implementation stages of the lifecycle, beginning with the system analysis phase, the obtained user requirements should be complemented with the requirements resulting from the law regulations concerning privacy and security. It especially applies to of data to be exchanged by Information systems in Supply Chain Management SCM in cloud environments as presented in [23]. Thus, the new technologies like cloud computing are offering potentially more secure data storage based on duplication and distribution [24].

In the system development and implementation stages there are some additional important issues to be regarded as crucial for more reliable protection of privacy and security aspects of software systems, such as:

- A deliberate choice of a system architecture,
- A secure (user) authentication procedure and data access control,
- A choice of the appropriate libraries according to the security requirements,
- The strict rules for the usage of external sources,
- The following the network security rules and the usage of appropriate computer network protocols,
- The proper packet management in mobile devices.

The first issue is a decision for a choice of the system architecture, like centralized, (or decentralized) or Cloud Computing usage, and it is dependent on the kind of a developed application, i.e. the usage of further technologies such as blockchain. The expansion of the systems based on new cloud technologies and cloud computing [25] by using services as IaaS – Infrastructure as the System, PaaS - Platform as a service, SaaS – System as a service, as also the growing usage of mobile devices set additional requirements on implementing solutions for an access to the remote data resources. It is particularly important to regard whether the system is processing personal data, according to the EU GDPR rules, or it is a transaction system – in accordance with the PSD2 directive. In such cases the additional considerations to cyber security are desirable and needed.

In case of the Client/Server n-tier architectures the sensitive personal data should reside on a back-end data server without a direct Internet access.

The second issue – the process of authentication procedure and data access control should be conducted in accordance with the requirements of PSD2 Directive. Therefore, we suggest to consider the two-factor authentication 2FA, with the cryptographic strong second factor. We can also recommend a usage of an authentication method based on Challenge-Response, as proposed in OCRA specification [26]. Moreover, the usage of one-time keys OTP is recommended, as mentioned in the previous Section. This way of usage of OCRA and OTP is simple to implement and will increase the security of the (user) authentication process.

The third implementation issue is a deliberate choice of appropriate libraries according to the security requirements. The libraries resistant to buffer overflow and the adequate programming methods for strict control of data types are recommended, as described in [12].

The next issue is a deliberate usage of external sources. The rules for conscious usage of external sources, especially for mobile applications should include the possible restrictions of a necessary usage of the external resources and libraries [28].

The network security rules and usage of appropriate computer network protocols should be carried out according to the principles indicated by W. Stallings and L. Brown in [12].

The last issue, is a proper packet management for mobile devices. A developed software should prefer the usage of packages that are resident on the device, i.e. not dynamically loaded during their usage at the run-time.

At the running (operational) phase, we also emphasize one more time the crucial role of a secure user authentication and the usage of dynamically changing one-time keys. These issues will be considered in the next two Sections.

#### IV. AUTHENTICATION METHODS

From the perspective of the enterprises the problems with data protection in business processes can be seen from the different perspectives, as described in [23] and [24]. The one perspective is considering the security and privacy of sensitive business data which belongs to the enterprise or its partners in the supply chain. Another point of view is the protection of the privacy of individuals.

Regarding the reliability and quality of service QoS for e-Business, the critical role of the IS security is one of the major business management responsibility. According to [4], the security encompasses the policies, the organizational procedures and technical measures used to guarantee a proper functioning of IS and protecting the enterprise against the consequences of malfunctioning. A guaranteed level of service performance should be delivered in

accordance with Service Level Agreement SLA addressing the QoS of the source [4].

Regarding the trust in e-Business, where the contacts related to transactions occur by means of databases and computer networks, there are additional trade risks like: man-in-the-middle attack, spoofing, hacking, denial of service attacks, etc. The extent of this kind of attacks covers the manipulation of data, intentional use of a false identity or attacking the enterprise's portal. Therefore the security requirements imply the infrastructure availability, the network level protection and the message security.

Moreover, the transparency and audit-ability of the transactions, their non-repudiation and certification are the further needs. The message (information) security requirements demand a safeguarding of a user authentication. In addition, information integrity and confidentiality are also crucial. In this paper we concentrate on one chosen aspect of the message security – the user authentication which is a combination of claiming an identity and its verification proving that the identity is as claimed [4].

For user authentication as a one of the message security requirements, a broad spectrum of the guidelines, and approaches, are known and available [12]. For instance, the rules for a user authentication are defined in the Digital Authentication Guideline NIST SP [27]. In this publication the authentication of the user is defined as the process of assurance (assertion) of the identity of the users introduced to the system. Further, in Protecting Controlled Unclassified Information in NIST [26] there are lists with security requirements for identification and authorization services classified as two basic requirements, and eleven secondary, derived requirements. A common digital model for digital authentication as defined in [27] includes few key roles entities and some functions needed in the authentication procedure.

In the above model, if the requester addresses the registration authority RA to become a subscriber of the confidential service provider CSP, the registration authority RA is the trusted party, that ensures (states and credits) the identity of the user (requester). The credential is the data structure connecting the identity with additional verifiable attributes needed in the process of authentication of the claimant to the verifier. In the process of verification of user authentication there are four basic common way like: some information known to the person, some kind of things possessed by the person (referred to as a token), the aspects to the physical person (the static or/and dynamic biometrics). Generally the enumerated factors can be used in separation, or can be combined together. Nevertheless, there could be some problems with using of each method. Therefore the multifactor authentication is considered to be a proper solution. In the next Section we present our approach to the authentication problem.

V. EXAMPLES FOR USER AUTHENTICATION - PROPOSED SOLUTIONS

Regarding the GDPR Regulation and PSD2 directives, from which the requirements considering privacy and information security should be fulfilled by the information systems, in this Section we will give some subsequent examples of a possible user authentication processes by using the HTTP protocol [29] for authentication aims for Web-based applications. The HTTP protocol is commonly used by the implementation of the Web-applications; at present the recommendation for business is the usage the HTTP/2.0+ push protocol.

In Fig. 1 are depicted the three stages of the gradual development of this protocol - beginning with HTTP/1.1, then the following HTTP/2.0, to the HTTP/2.0+ push development stage. The development process of the subsequent stages has taken into account the aim of the minimization of the needed connections between the client and the server of a software application.

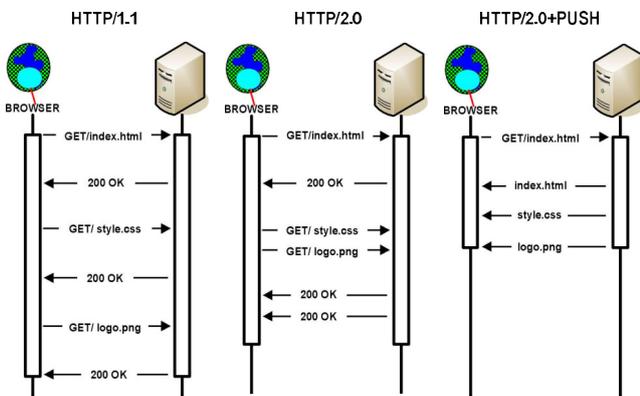


Fig. 1 Evaluation stages of HTTP protocol development

Using HTTP/1.1 protocol for opening of a Web page requires even three full transactions with the transmission of the elements containing the website descriptions. However, in the currently recommended version by W3C of the HTTP/2.0+ push protocol, the opening or refreshing of a website requires only one connection. The eliminating of unnecessary connections also significantly positive affects the security concerns, because it eliminates at the same time the possibility of seizing sessions and man-in-the-middle attacks to the necessary minimum, and this way reduces the risk of impersonating another users. It seems right, that good practices used in the development of HTTP, should also be utilized in systems applied for user authentication.

The second example is a simple authentication mechanism shown in Fig. 2, which uses only one single connection. This authentication method uses the asymmetric cryptography, also known as public key cryptography, which means, it uses public and private keys to encrypt and decrypt data. But in the process of the authentication the Client is using a secure

private key denoted as  $prK$  and a secret password  $P$  to compute a cryptogram:

$$\text{Cryptogram } S = \text{Crypt}(P, prK).$$

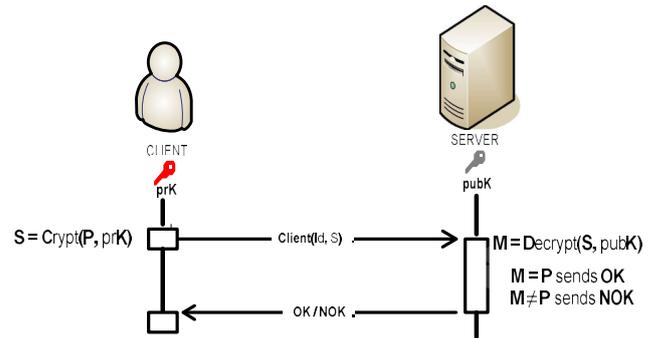


Fig. 2 Example of a simple authentication using asymmetric cryptosystems

The server, which knows the shared secret password  $P$ , performs the decryption by using the public key  $pubK$  in the method  $\text{Decrypt}(S, pubK)$  and verifies if the shared secret is known by the client; the cryptographic function  $\text{Crypt}()$  is a known cryptographic algorithm with a public key. As a cryptographic algorithm proposed for the analysis aims we recommend the usage of the RSA schema described in [30]. In this method, as shown in Fig. 2, the authentication mechanism uses only a single connection.

However, this simple authentication method can not be regarded as secure, and is vulnerable to the various attacks by unauthorized users. A simple attack is presented in Fig. 3, where the hacker has an option of intercepting the cipher text  $S$  and so impersonating another user by using the captured cipher text. This method is referred to as sniffing accomplished by the Web communications and spoofing by another user.

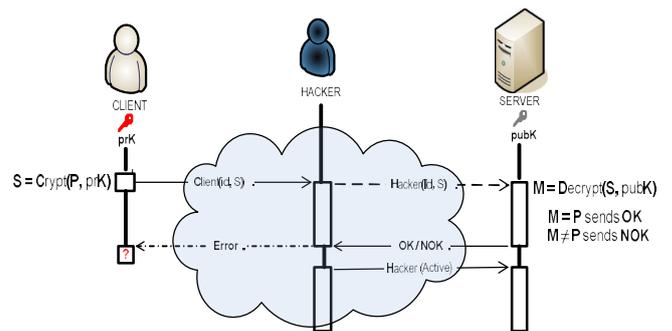


Fig. 3 The simple attack at authentication

The most secure authentication method is based on asymmetric cryptography and a cryptographic hash function named Hash [31]. This is a mathematical algorithm that maps data of an arbitrary size to a bit string of a fixed size (a hash) and is designed to be a one-way function, that is, a function, which is infeasible to be inverted [31]. The

proposed method uses the asymmetric cryptography and the cryptographic hash function - SHA2, in the authentication case shown in Fig. 4.

One must realize that removing parts of data will lead to results with lesser granularity, but this is a price, which must be paid to stay on the safe side and to stay compliant with the privacy rights.

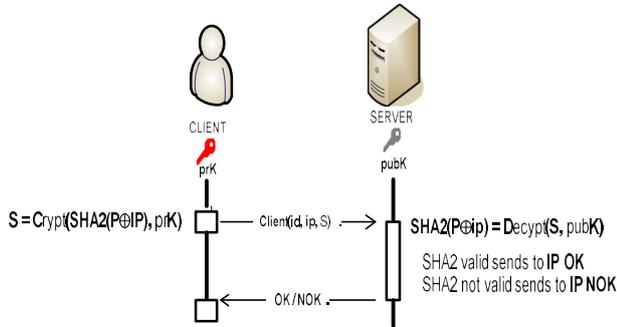


Fig. 4 Example authentication using asymmetric cryptosystems

In contrast to the known methods, which are using the hash function, in this solution the *IP* – the Internet protocol number was proposed as one of the input parameters for the hash function  $SHA2(P \oplus IP)$ . In this case the secret password *P* is aggregated with known *IP* and from this value a Client generates a cryptogram  $Crypt(SHA2(), prK)$ . The proposed method is eliminating the possibility of simple spoofing, because this way we authenticate only the user, which has the valid *IP* number. However, acting according to this scheme does not protect against cryptanalytic attacks known to asymmetric cryptographic systems such as RSA.

In the scheme depicted in Fig. 4 we use a fixed cryptographic key, and this way we allow the off-line analysis of the private cryptographic keys.

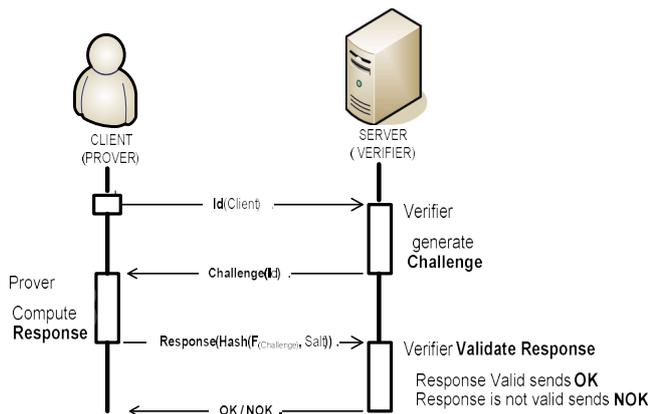


Fig. 5 Challenge-Response method in Client authentication method

In Fig. 5 there is a new situation depicted, where a Client uses another authentication method named the Challenge-Response protocol. In this method only the hash functions

are used; the method is used in OCRA. This method uses a constant function for a validation of the Response, and the Challenge can be regarded as variable value changing with each authentication.

In the last example (see Fig. 6) we give a proposal of Challenge-Response method based on the solution described in [20].

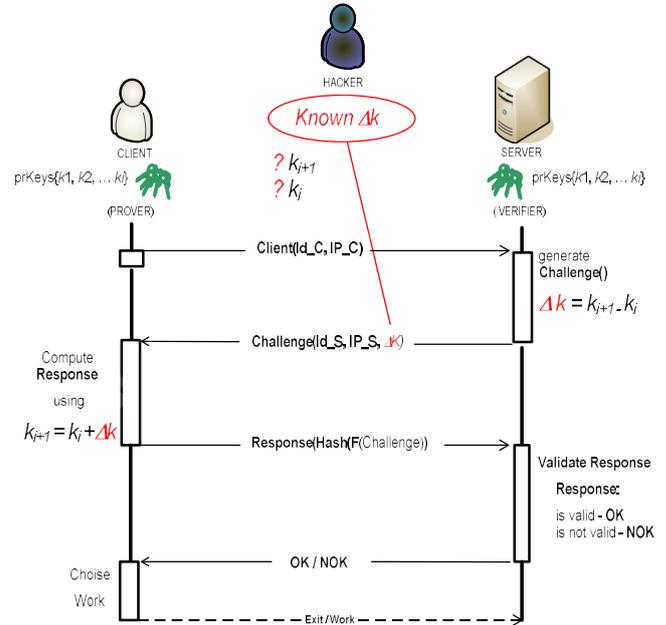


Fig. 6 Proposed Challenge-Response with OTP method in Client authentication

In Fig 6. a Challenge-Response authentication method is depicted, which is offering the perfect security of an authentication. In this solution, a high level of security is achieved through the authentication, which is using the advantages of OTP in asymmetric cryptography encryption and the hashing SHA2 method in the Challenge-Response mechanism. Additionally this proposal uses RSA, where the changeable shared value *prKeys* is a set of  $\{k_1, k_2, \dots, k_i\}$  and as a Challenge we are using the differential value  $\Delta k = k_{i+1} - k_i$ . In this case, even if someone is knowing the value  $\Delta k$ , it will be not possible to determine either the  $k_i$  or  $k_{i+1}$ .

The last shown method guarantees the semantic security, i.e. an adversary hacker can not gain even a partial information about an encrypted message (password) [32]. Therefore we recommend to undertake the user authentication in information systems collaborating in supply chains with usage of this method, as being the most appropriate for satisfying the requirements of cyber security, due to the EU regulations.

The last solution presented in this Section (depicted in Fig. 6) also conforms to the requirements of OTP and should be also resistant to the attacks with quantum computing, because, as stated above, the equation system given in the Fig 6 has no solution.

In this Section we gave some concrete examples for the authentication of the user (or processes) and also a proposal for an authentication of an information source, meeting the requirements ISO/IEC 2911, and also the European Directive from the 8th of September 2015 No. 1502 regarding the dynamically authentication.

## VI. CONCLUSION

Already Alvin Toffler, Future Shock and Third Wave author, has indicated that the next phase of the industrial development will be the information society, where the information will have a particular value [33].

In the context of a rapid development of modern information technologies and digitalization, there is a growing importance of new factors that could threaten the security of logistics processes. It may be due to the wrong decisions caused by false (unreliable, insufficient, or incorrect) information, or caused by non-compliance with the required procedures or wrong (false) documents.

A solution trying to tackle such vulnerability problems is for instance a proposal of the COBIT 5 Information Security Framework for reducing cyber attacks on SCM systems [34]. However the COBIT framework does not take into consideration the concerns associated with the security of the authentication, which are based on the norms ISO/IEC 29115, and the European directive 1502/2015, as nowadays required.

According to [35] there are seven security concerns, which are addressing the main problems in contemporary supply chains: the inventory theft, the mismanagement of cloud access, the smuggling, the increased piracy, the physical device tampering, trusting data to a third party vendor, and the IoT Sensor compromise.

The physical threats like the smuggling or the piracy are out of scope of the considerations of this paper. The inventory theft is in fact a physical threat, however the credibility of the stocks of the inventories still remains important. The physical device tampering can cause the corruption of the data or disruption of the devices (or chips). Furthermore, the IoT sensor data represents an another possible attack vector.

From the above concerns particularly the mismanagement of cloud access due to an improper authentication could lead to serious security risks for supply chains. Therefore credibility of data, and also of the data sources, and on the other hand the audit-ability of the vulnerabilities of enterprise system are becoming crucial to guarantee the security in SCM.

Another important aspect, due to the GDPR is the preventing of data privacy for enterprises offering their services in the EU. In [36] the authors propose an approach for deriving Workflow Privacy Patterns from legal texts; these patterns are meant to support the designing of privacy compliant workflows.

In our paper we have shown some solutions as a proposal for complementing the tools with the elements of a secure authentication, that are also applicable in the context of a usage of the blockchain technology.

The design of software applications with respect to the demanded security and privacy requirements remains one of the current challenge for development and usage of Information systems. The new EU Regulation GDPR and PSD2 concerning the technologies used to support the banking and financial services Fintech, draw attention to the enhancement of required assurance level in security for processing sensitive data. The increasing numbers of incidents with the data leaks and an unauthorized access to digital resources or the denial of service (DOS) and other attacks are the symptoms of the raising problems with proper dealing with cyber security of the systems.

As a one most weak constituent in the system security considerations is the human factor which can not be so easily eliminated. In this paper we suggest the usage of the proper mechanisms, methods, and technologies that could be involved into the life cycle of are the Information systems, (as the constituents of supply chains) with the aim to increase the security of data and the transactions. Accordingly we have highlighted the importance of the deliberate choice of a software architecture, following the security rules during data exchange via computer networks, and also recommend a usage of the technologies, which are viable for a secure user authentication i.e., those with one-time keys.

In the examples in the last Section we have shown some possibilities for reducing the number of the required connections during the user authorization process in order to reduce a possibility of the hacker attacks.

The improvement of information security for information systems, can be achieved especially by using carefully chosen user authentication methods with the aim of fulfillment of the requirements of the high assurance levels of EU regulations.

The aim of this paper was to present the authentication methods conforming to the all three assurance levels given in [16]. The proposed solution presented in Section 5, in the last sixth example is a unique new solution, which fulfills the substantial and high assurance level of this EU Directive. The proposed method guarantees the realization of a dynamic authentication needs, as defined in this regulation and required by GDPR and PSD2. Thus we suggest the usage of this method in association with the technologies like block chain and cloud computing.

In the future we will further investigate the security aspects of Information systems and especially consider the diverse methods for a secure user authentication.

Currently recommended is the usage of HTTPS and TLS protocols with elliptic curve sieve [37] with a small size of the encryption key. In the future we will consider the enhancement of this solution where the usage of RSA in the cryptography could be substituted by the application of the

elliptic curves or a lattice-based cryptography. Such solutions will be needed with the emerging quantum computing (post-quantum cryptography).

#### REFERENCES

- [1] General Data Protection Regulation, “Regulation on the protection of natural persons with regard to the processing of personal data and on the free movement of such data” EU 2016/679, 2016.
- [2] Payment Services Directive 2, “Directive on payment services in the internal market”, EU 2015/2366 Official Journal of the European Union Payment Service Directives 2. EU 2015/2366, 2015.
- [3] Fintech: [www.investopedia.com/terms/f/fintech.asp](http://www.investopedia.com/terms/f/fintech.asp)
- [4] M. P. Papazoglou, and P. M.A. Ribbes, *E-business: organizational and technical foundations*, John Wiley and sons. London, 2006.
- [5] L. Gil, and A. Liska, “Security with AI and machine learning. using advanced tools to improve security at the edge”, New York O’Reilly, 2019.
- [6] Global Economic Crime and Fraud Survey, Pulling fraud out of the shadows. The biggest competitor you didn’t know you had. 2018. <https://www.pwc.com/gx/en/forensics/global-economic-crime-and-fraud-survey-2018.pdf>
- [7] D. Bohn, “Android at 10: the world most dominant technology”, 2018 <https://www.theverge.com/2018/9/26/17903788/google-android-history-dominance-marketshare-apple>
- [8] Hybrid warfare. Wikipedia [https://en.wikipedia.org/wiki/Hybrid\\_warfare](https://en.wikipedia.org/wiki/Hybrid_warfare)
- [9] T. Magee, “The most notorious hacker groups”, ComputerworldUK <https://www.computerworlduk.com/security/most-notorious-hacker-groups-3679258/>
- [10] G. Perkovitz and A. E. Levite, Eds., “Understanding Cyber Conflict”, Georgetown University Press, 2017.
- [11] D. Sorin, The cyber dimension of modern hybrid warfare and its relevance for NATO Europolitics, vol. 10-1, 2016. <http://europolity.eu/wp-content/uploads/2016/07/Vol.-10.-No.-1.-2016-editat.7-23.pdf>
- [12] W. Stallings, and L. Brown, “Computer Security: Principles and Practice”, Pearson Education 2018.
- [13] OWASP Foundation. The free and open software security community, <https://www.owasp.org>
- [14] C. Wang, Q. Wang, K. Ren, N. Cao, and W. Lou, “Toward secure and dependable storage services in cloud computing”, IEEE Transactions on Services Computing, vol. 5-2, April-June 2012, pp. 220 – 232, DOI: 10.1109/TSC.2011.24
- [15] D. Mills, K. Wang, B. Malone, A. Ravi, J. Marquardt, Chen, A. Badev, T. Brezinski, L. Fahy, K. Liao, V. Kargenian, M. Ellithorpe, W. Ng, and M. Baird, “Distributed ledger technology in payments, clearing, and settlement”, Finance and Economics Discussion Series 2016-095, 2016. Washington: Board of Governors of the Federal Reserve System, <https://doi.org/10.17016/FEDS.2016.095>.
- [16] Official Journal of the European Union. Technical Specification for assurance levels for electronic identification. 1502/2015EN.
- [17] C. E. Shannon, “Communication theory of secrecy systems”, The Bell System Technical Journal, vol. 28-4, Oct. 1949.
- [18] L. Chen, S. Jordan, Y-K. Liu, D. Moody, R. Peralta, R. Perlner, and D. Smith-Tone, “NISTIR 8105 Report on Post-Quantum Cryptography”, <http://dx.doi.org/10.6028/NIST.IR.8105> <https://nvlpubs.nist.gov/nistpubs/ir/2016/nist.ir.8105>
- [19] Adips, RUBLON, “Trusted access multi-factor authentication”, Zielona Góra, 2016. <https://rublon.com/>
- [20] J. Jabłoński, “Encryption system with one-off key”, no. 218339, submitted 20-04-2011, date of the patent 10-09-2014.
- [21] Project POIR .01.01.01-00-0257/16 - CryptOne unconditional secure crypto-processor, DCD Digital Core Design Bytom, Poland 2016-2019.
- [22] J. Jasperneite, “What is Industrie 4.0”, Computer&Automation, 2012
- [23] S. Robak, B. Franczyk, and M. Robak, “Business process optimization with big data analytics under consideration of privacy”, Proceedings of the 2016 Federated Conference on Computer Science and Information Systems, M. Ganzha, L. Maciaszek, M. Paprzycki (eds). ACSIS, Vol. 8, p 1199–1204, 2016, DOI: <http://dx.doi.org/10.15439/2016F542>
- [24] B. Schwarzbach, M. Glöckner, A. Pirogov, M. M. Röhling, and B. Franczyk, “Secure service interaction for collaborative business processes in the inter-cloud,” in 2015 Federated Conference on Computer Science and Information Systems, ser. Annals of Computer Science and Information Systems, IEEE, 2015, pp. 1377–1386. <http://dx.doi.org/10.15439/2015F282>
- [25] D. Agrawal, S. Das and A. E. Abbadi, „Big data and cloud computing: current state and future opportunities“. EDBT 2011, March 22-24, 2011, Uppsala, Sweden. ACM 978-1-4503-0528-0/11/0003.
- [26] RFC 6287 “OCRA: OATH Challenge-response algorithm”, Internet Engineering Task Force IETF 2011, <https://tools.ietf.org/html/rfc6287>
- [27] P. Grassi, M. Garcia, and J. Fenton, “Digital authentication guideline”, NIST SP 800-63-3, 2016.
- [28] R. Ross, K. Dempsey, P. Viscuso, M. Riddle, and G. Guissanie, “Protecting controlled unclassified information in nonfederal information systems and organizations” NIST SP 800-171, 2016.
- [29] HTTP - Hypertext Transfer Protocol, <https://www.w3.org/Protocols/>
- [30] S. Rivest, A. Shamir, and L. Adleman, “A method for obtaining digital signatures and public-key cryptosystems”, Comm. of the ACM, vol. 21-2, 1978, pp. 120–126.
- [31] B. Schneier, “Cryptanalysis of MD5 and SHA: Time for a new standard”, Computerworld, 2014.
- [32] S. Goldwasser, and S. Micali, “Probabilistic encryption”, Journal of Computer and System Sciences, vol. 28-2, 1984, pp. 270-299. [https://doi.org/10.1016/0022-0000\(84\)90070-9](https://doi.org/10.1016/0022-0000(84)90070-9)
- [33] A. Toffler, *The third wave*. Bantam Books, 1980.
- [34] M. Wolden, R. Valverde, and M. Talla, “The effectiveness of COBIT5 information security framework for reducing cyber attacks on supply chain management system”, IFAC-PapersOnLine Vol. 48-3, 2015, pp. 1846-1852. <https://doi.org/10.1016/j.ifacol.2015.06.355>
- [35] L. Wainstein, “7 supply chain security concerns to address in 2019”. <https://supplychainbeyond.com/7-supply-chain-security-concerns-to-address-in-2019/>
- [36] M. Robak, and E. Buchmann, “Deriving workflow privacy patterns from legal documents”, Federated Conference on Computer Science and Information Systems, 2019 – accepted paper.
- [37] V. Gupta, D. Stebila, S. Fung, S.C. Shanz, N. Gura, and H. Eberle, “Speeding up Secure Web Transactions Using Elliptic Curve Cryptography”, <http://research.sun.com/projects/crypto>

# Deriving Workflow Privacy Patterns from Legal Documents

Marcin Robak

Hochschule für Telekommunikation Leipzig, Germany  
Email: robak@hft-leipzig.de

Erik Buchmann

Hochschule für Telekommunikation Leipzig, Germany  
Email: buchmann@hft-leipzig.de

**Abstract**—The General Data Protection Regulation (GDPR) has strengthened the importance of data privacy and protection for enterprises offering their services in the EU. An important part of intensified efforts towards better privacy protection is enterprise workflow (re)design. In particular, the GDPR as strengthen the imperative to apply the *privacy by design* principle when (re)designing workflows. A conforming and promising approach is to model privacy relevant workflow fragments as Workflow Privacy Patterns (WPPs). Such WPPs allow to specify abstract templates for recurring data-privacy problems in workflows. Thus, WPPs are intended to support workflow engineers, auditors and privacy officers by providing pre-validated patterns that comply with existing data privacy regulations. However, it is unclear yet how to obtain WPPs systematically with an appropriate level of detail.

In this paper, we introduce our approach to derive WPPs from legal texts and similar normative regulations. We propose a structure of a WPP, which we derive from pattern approaches from other research areas. We also introduce a framework that allows to design WPPs which make legal regulations accessible for persons who do not possess in-depth legal expertise. We have applied our approach to different articles of the GDPR, and we have obtained evidence that we can transfer legal text into a structured WPP representation. If a workflow correctly implements a WPP that has been designed that way, the workflow automatically complies to the respective fragment of the underlying legal text.

## I. INTRODUCTION

**P**RIVACY and data protection are within the scope of interest of enterprises since years. Most current privacy related efforts in enterprises are driven by the General Data Protection Regulation (GDPR) [1] which came into action in May 2018 at the EU level. The regulation describes a set of imperatives enterprises have to consider in their workflows. A workflow is a business process automation, where information and tasks are transferred between participants according to business rules. Regarding GDPR, special attention should be paid to the Article 25 ('data protection by design and by default'). It obliges businesses to implement privacy-aware data management processes in all workflows that handle personal data. This is a complex and challenging task, because all respective workflows must be reconsidered from a privacy perspective. These requirements can originate from privacy norms written in national and international law texts. They also can result from a company's Binding Corporate Rules.

Workflow Privacy Patterns (WPPs) have been introduced by [2]. The idea of WPPs is to compile complex data privacy norms into a compact representation which support workflow

creators and analysts with designing and verifying workflows. WPP have to be pre-validated by data privacy experts and must be understandable for a wider audience. Workflow engineers without legal expertise shall be able to assess if the implementation of a particular WPP allows to create a privacy-compliant workflow. The implementation of a WPP shall not require legal expertise. Also, it shall be easier for a workflow analyst to find out if a workflow contains a WPP, than to conduct a privacy assessment unassisted. Thus, the WPP approach is promising. However, what is currently missing is a library of validated WPP designs. This is due to the fact that there is no approach to obtain WPPs from legal sources. In this paper, we introduce our approach to derive WPPs from complex legal texts containing data privacy norms.

Our research method is based on the design science [3] approach. We start with a problem statement, then we systematically compile a set of requirements for 'good' WPPs. Based on the structure of legal documents, we deduce which information must be represented in a WPP, and we provide a framework to extract this information from documents such as binding corporate rules, national and international law texts or compliance rules. We show applicability of our approach with two different use cases.

Our work indicates that it is possible to create WPPs in a structured way, resulting in WPPs with practical potential. This could foster companies in fulfilling privacy obligations which promote customer privacy protection.

**Paper structure:** The next section describes fundamentals and legal concepts related to our work and serves as a starting point for our research. In Section III we define a structure of a WPP, and in the Section IV we describe how to fill it with content derived from legal documents. This section also shows exemplarily how this framework can be applied to a fragment from the GDPR. Finally, Section V concludes.

## II. RELATED WORK

In this section we discuss legal and research foundation related to data privacy. We will also describe the concept of patterns which is in use in the computer science and other industry areas.

### A. Privacy concepts

The GDPR describes several requirements on privacy; most of them are well-proven concepts. The GDPR has an impact

on workflow designs on three different levels of abstraction:

On a global level, the GDPR obligates the enterprises to take care about data protection already *while planning and designing* their workflows. Specifically, Article 25 requires that the processing of personal data shall be planned and executed always in a way which supports privacy. This requirement is known also as privacy (or data protection) by design and by default [4]. It results from postulate of instant protection, and from the observation that effective data protection should not be realized only by reactive or retrospective actions [5]. To obtain privacy by design, other two levels must be taken care of. We describe them below.

The second level of the GDPR's impact on workflows is the *requirement for particular actions* in specific situations. Several Articles describe situations for which particular actions must be taken. For example, Article 15 ('right of access') calls for businesses that provide information about the amount of personal data, the purposes of the processing, its storage period, etc., as soon as a person files a request for information. Other articles describe further situations the enterprises must be prepared for. It can be changing or erasing personal data, if a person asks for it in line with the Article 16 ('right for rectification') or Article 17 ('right to be forgotten').

The third level is constituted by the *principles* relating to the processing of personal data. They do not describe specific actions or workflow fragments, but they still affect workflows. Some of these principles are described in the Article 5. For example 'purpose limitation' principle requires that the data collected to fulfill one particular business task should not be used for other purposes. The data minimization principle specifies that the amount of personal data which is collected or handled should be limited to the minimum required to finish the business task.

### B. Patterns

Design patterns are reusable solutions for recurring problems. Design patterns have been proposed in several fields. Already in 1977 Alexander [6] wrote "Each pattern describes a problem which occurs over and over again in our environment, and then describes the core of the solution of the problem, in such way that you can use this solution a million times over, without ever doing it the same way twice". The same kind of thinking was adapted in the fields of software engineering [7] and IT architecture [8].

In the field of workflow modeling, workflow patterns have been introduced [9]. Different perspectives of workflow models can be considered [10], depending on the intended use of the model. Well-known perspectives are 'control flow', 'data', 'resources', 'functional' and 'operational'. Most workflow patterns [11] focus on the first three perspectives. For example, [12] lists 43 different control-flow patterns ranging from the synchronization of parallel workflows to the explicit termination of workflows. Patterns regarding the data perspective [13] consider the visibility of data, data-driven interactions, the transfer of data and its transfer routes. Patterns such as 'Role-based allocation' [14] address the life cycle of work items

from the resources perspective. [15], [16] present exception handling patterns.

In the area of data privacy, collections of software design patterns have been already proposed [17], [18]. Such collections include options to collect, process and share personal data in a legal way, e.g., by using anonymization, onion routing or implied consent. However, a structured collection of design patterns for the data-privacy perspective in workflows does not exist so far.

### C. Representation of privacy requirements

In general, three approaches exist to integrate privacy requirements into workflows. They vary in the degree of abstraction and the degree of formalization.

Numerous 'best practice' *implementation guides* have been written by privacy authorities, privacy officers and law firms. Such guides contain textual descriptions of steps needed to handle legal obligations. For example, a guide could translate a GDPR Article into an intuitive description of steps which have to be performed. In many cases the guides are tailored to specific industry sectors. However, such guides are less structured than the legal articles. This induces some degree of freedom when implementing them into workflows. Thus, it is difficult to ensure that a workflow designed on basis of a guide is indeed compliant with the regulation.

*Checklists* allow to perform a target-actual comparison in a structured way. A checklist reduces the effort needed to incorporate legal requirements into workflows. A legal article is distilled to a list of capabilities which must be implemented. However, it is difficult to express some legal obligations only in form of one-dimensional checklists. For example, it would be confusing to represent the right of access as a checklist. This is because the right of access is interwoven with other articles of the GDPR, depending on aspects such as data transfers into third countries or conflicts with the rights of other persons.

Finally, industry-specific *reference models* provide optimized workflow models in a semi-formal language such as EPC [19] that handle typical privacy obligations. For example, a domain expert could define a reference model for handling incoming requests for access in a typical retailer scenario. Thus, the reference model contains best practices in a specific application domain. A workflow engineer could adapt this model to the workflows of his company. However, a reference model does not ensure that its implementation into the workflows of a company is correct regarding the privacy obligation. This has two reasons: Firstly, languages such as EPC or BPMN do not allow to model all obligations mentioned in privacy regulations, e.g., storage periods or data transfers to foreign countries with less developed privacy standards. Secondly, the workflow engineer has a high degree of freedom when adapting the reference model to his company.

## III. DERIVING WORKFLOW PRIVACY PATTERNS

Workflow models automate business processes that execute specific business tasks. To design a workflow model, a workflow engineer analyzes business objectives, company structure,

key performance indicators, etc. But also legal obligations must be met. This is where data privacy requirements come into play. They have an impact on workflow design and are involved in several aspects of workflows. For example, the order of activities (the sequence flow order) in a workflow is vital for privacy. A natural person must give consent *first*, before his data is stored or processed. The data flow within workflows is another important aspect. Authorization and authentication for gaining data access must be carefully planned. Also execution exceptions have the potential to violate data privacy regulations, say, if an activity on personal data cannot be completed without involving third parties.

Consider Text 1, which we will use as a running example in this paper. It shows a typical article from the GDPR.

*Text 1 (Fragment of GDPR's Article 15 - Right of access):*

1. *The data subject shall have the right to obtain from the controller confirmation as to whether or not personal data concerning him or her are being processed, and, where that is the case, access to the personal data and the following information:*
  - (a) *the purposes of the processing;*
  - (b) *the categories of personal data concerned;*
  - (c) *the recipients or categories of recipient to whom the personal data have been or will be disclosed, in particular recipients in third countries or international organisations;*
  - (d) *where possible, the envisaged period for which the personal data will be stored, or, if not possible, the criteria used to determine that period;*
  - (e) *the existence of the right to request from the controller rectification or erasure of personal data or restriction of processing of personal data concerning the data subject or to object to such processing;*
  - (f) *the right to lodge a complaint with a supervisory authority;*
  - (...)
2. *Where personal data are transferred to a third country or to an international organization, the data subject shall have the right to be informed of the appropriate safeguards pursuant to Article 46 relating to the transfer.*
3. *The controller shall provide a copy of the personal data undergoing processing. (...)*

Staying compliant with such legal regulations implies many consequences for a company's workflows. Enterprises must be prepared for the case when a customer places such access enquiry and they must be able to react accordingly.

#### A. Problem Statement

A WPP is a translation of one or more privacy obligations into a semi-formal specification, which can be integrated into a workflow model [2]. WPPs support enterprises to be compliant with data privacy regulations. In particular, WPPs shall foster

planning, implementing and auditing of workflows handling personal data. In order to find out how such a WPP must be structured and how it can be obtained in a systematic way, we need to consider the capabilities of the WPP users, and we need to define requirements that a WPP must fulfill in order to be applicable.

*a) User roles:* We have analyzed which different roles are involved in creation and use of WPPs. Our focus was on the functions the roles must fulfill, and which knowledge and which skills are needed in this regard. We have identified three distinct user roles:

**WPP creator** This role develops a WPP from a particular data privacy norm. This role has legal expertise needed to identify all information from various legal sources, that must be considered in order to implement privacy-compliant workflows. This skill is needed to be able to mirror the legal norm(s) semantically. The WPP creator needs background knowledge on workflow modeling to provide syntactically correct WPPs.

**Workflow engineer** This role models workflows with the help of WPPs. The workflow engineer implements WPPs into existing workflows or creates new workflows according to a WPP specification. This role needs domain knowledge on the workflow domain and workflow modeling skills, but it doesn't need to possess legal knowledge.

**Privacy officer** This role verifies and documents if workflows are compliant with data privacy norms. In this role can be a employee or an external auditor. A privacy officer has sufficient domain knowledge and legal expertise to find out, if existing workflow model meets certain privacy obligation.

*b) Requirements for WPPs:* From the intended use of the WPPs and the expertise of the user roles, we have derived three requirements for WPPs:

**R1** WPPs are a variant of design patterns. Thus, WPPs have to meet all *general requirements for design patterns*, e.g. completeness, understandability and reusability.

**R2** Because the workflow engineer may lack legal expertise, a WPP must contain *all information necessary* to model or validate a certain privacy obligation. For example, if a WPP is a specification for the implementation of the 'right of access' - as shown in Text 1 -, then it must be possible to create a privacy-compliant workflow on the basis of this WPP only, i.e., without having to consider additional legal texts.

**R3** WPPs must be modular to enable *linking of WPPs*. This is particularly important, as privacy obligations often are spread over several articles or multiple legal texts.

Given these requirements, we will now explore options to structure WPPs. We start by deriving an information model to express information from legal norms in a WPP. In the next section, we propose our framework to compile WPPs from legal texts.

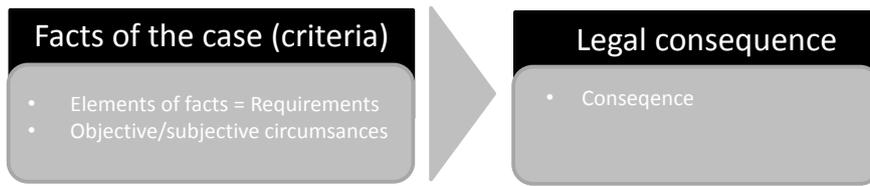


Fig. 1. Structure of legal texts

*B. Structures of legal texts and design patterns*

In this subsection we compare the structures of legal texts and design patterns. Obligations in legal texts typically follow a well-defined structure, as shown in Figure 1. A legal obligation is described by

- (1) **the facts of the case** and
- (2) **the legal consequences.**

The facts of the case specify

- (1a) the *general criteria* for the applicability of the norm and
- (1b) the *circumstances* under which a certain legal norm shall be applied.

The facts of the case result in an if-then form. Thus, the legal norm or corporate rule can be always interpreted as 'if all prerequisites are met, then the consequences apply'. The consequences in turn can be either

- (2a) a *course of action* that must be taken or
- (2b) a *yes/no-conclusion* in the sense 'if all prerequisites are met, then the regulated action is lawful'.

In a case of our running example, the general criteria for the applicability of the norm (1a) are described in Art. 2, 3 GDPR (Text 2, 3). The norm applies if the company handles personal data related to activities in the EU.

*Text 2 (Fragment of Article 2 GDPR): Material Scope*

1. This Regulation applies to the processing of personal data wholly or partly by automated means and to the processing other than by automated means of personal data which(...)

*Text 3 (Fragment of Article 3 GDPR): Territorial Scope*

1. This Regulation applies to the processing of personal data in the context of the activities of an establishment of a controller or a processor in the Union, (...)

The circumstances (1b) for a person claiming access rights are described in the first paragraph of Art. 15 GDPR (Text 1). It says that the company must actually possess information about this person. The legal consequence (2) is described in the subsequent paragraphs of Art. 15. The consequence requires the company to provide certain information (2a), according to further dependencies.

Design patterns consist of three components, as described in the previous subsection: (i) the *context* the pattern can be applied to, (ii) the *problem* description that allows the engineer to decide, if the pattern is useful for specific design problem, and (iii) a generic *solution* for the described problem [20]. Observe that the general structure of design patterns is similar to the structure of obligations in legal texts; this is shown in the Figure 2. Thus, it seems appropriate to define a WPP alike. To this end, we distinguish **activity patterns** where the consequence is a course of action (2a), and **check pattern** that result in a yes/no-conclusion (2b).

*C. Options to represent legal texts*

In order to obtain evidence on approaches to structure a WPP, we have conducted a series of preliminary experiments.

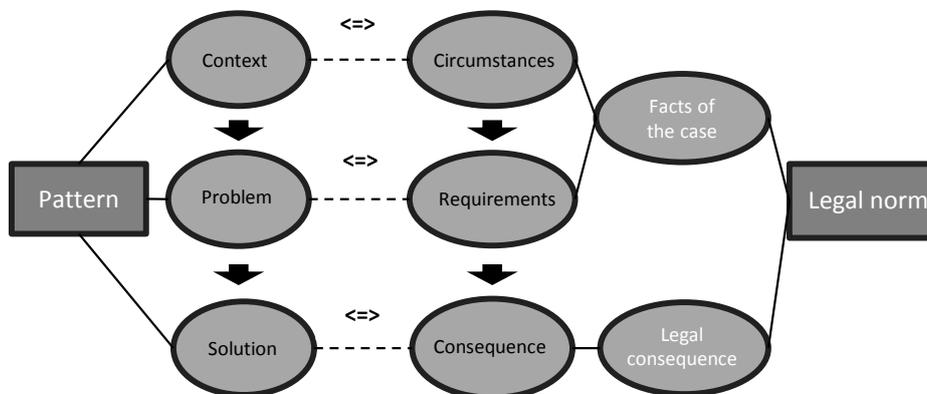


Fig. 2. Relation between legal texts and design patterns

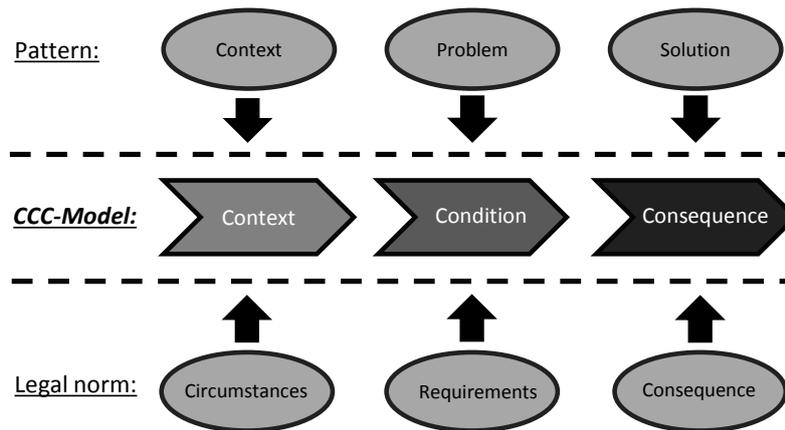


Fig. 3. CCC-Model

In particular, we have asked a class of master's students to model the facts of the case and the legal consequences of various articles of the GDPR. The students had a professional background on data privacy and security and attended an extra-occupational education class on workflow modeling.

The students have observed that the general criteria for the applicability of the norm (1a) refer to domain knowledge of the workflow that cannot be easily represented as a check list or a BPMN-style workflow model. We think that describing the criteria textually is the most appropriate option. Furthermore, our students have reported that the set of circumstances for the applicability of a specific article (1b) does not have an inherent order. Therefore it makes no sense to represent the circumstances as a workflow model fragment with a graphical language. A simple check list is sufficient and was preferred by the students. Our students also found out, that the legal consequence (2a) can be represented as workflow model. This model can be defined in a semi-formal language such as BPMN or EPC. If the consequence is a straightforward yes/no-conclusion (2b), this part can be cut down to a simple event 'Processing is lawful'. The final observation of the experiment was, that only such articles can be represented in a proposed way, which do not contain uncertain legal concepts. For example, consider Text 4. It requires legal expertise to decide for each workflow instance individually if the interests of the controller are overridden by the rights of a person.

*Text 4 (Frag. of Art. 6 GDPR): Lawfulness of processing*

1. Processing shall be lawful only if (...)
  - (f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject (...)

#### IV. THE CCC MODEL

In this section, we introduce our CCC model. It structures fragments of legal texts into Context, Condition and Consequence. In the previous section, we have observed a similarity between the general structure of legal texts (the circumstances for applicability of an article, the legal requirements named in the article and the legal consequence) and three basic elements of a pattern (context, problem, solution). This similarity is outlined in Figure 3. Furthermore, we have obtained evidence how the different parts of legal texts can be represented. In this section, we first describe elements of a WPP structure, and then follow up with the CCC Model, which describes how to obtain systematically WPPs from legal texts.

##### A. WPP structure

We aimed for WPP structure elements that mirror and foster desirable characteristics of design patterns, such as completeness, understandability and reusability (Requirement R1). Furthermore, the structure of a WPP shall carry all legal obligations from the data privacy domain for a given scope (Requirement R2). It shall not result in oversized, inapplicable pattern forms, that violate the Requirement R1. The structure must allow modular stacking of WPPs (Requirement R3). Considering this, our WPP structure consists of *Header*, *Context*, *Condition* and *Consequence*:

a) *Header*: The header contains meta-information of the pattern. It describes essentials like name, type, legal focus of the WPP and relation to other WPPs. Further meta-information as an unique database ID, date or the name of the WPP creator, may be added.

**WPP Name** A distinct name of the pattern. It makes the pattern easily recognizable, and allows searching for it in a pattern catalog. The name of the WPP shall indicate the objective of the pattern.

**WPP Type** WPPs can be distinguished into check patterns and activity patterns, as observed in the last section.

**Legal Focus** Specifies all legal texts (articles, paragraphs, etc.) which were used to derive the pattern. It declares which legal obligation is covered (entirely or partly) with this WPP.

**Relation to other WPPs** WPPs can build upon each other. When implementing multiple WPPs into a workflow, sometimes the relation between WPPs needs to be specified. For example, a WPP creator might decide to split the legal obligation to delete data into multiple WPPs. One WPP keeps records of the data used, a second one ensures that the data is deleted at the specified time. WPPs might also exclude each other. For example, a WPP to execute a business task anonymously might exclude a WPP for the deletion of personal data. Since new WPPs might be created at any time, information on the relation to other WPPs may be incomplete.

b) *Context*: The context of a WPP contains an intuitive textual description of the situation and of the resulting problem, which is addressed by the WPP. The user must clearly understand when and for which objectives the WPP can be applied, and if the application of the WPP results in further legal obligations.

c) *Condition*: The condition provides all prerequisites mentioned in the legal texts that have been enumerated in the 'legal focus' field of the WPP header. Since the order of the prerequisites is insignificant, the condition is represented as a checklist. The prerequisites have to be defined as positive statements that do not leave room for misunderstanding. If all prerequisites in the checklist are met, the consequence applies.

d) *Consequence*: The consequence of a check pattern is a statement, which is true, if all prerequisites from the condition are fulfilled. In order to determine the consequence of the WPP, it is necessary to specify the type of the pattern first. This is, because the consequence component differs in its form depending on the type of the WPP. For a check pattern it is (a) a statement that the case described in the context field is lawful, according to the legal norms specified in the header. Alternatively - for an activity pattern - the consequence is (b) a chain of activities, specified with a workflow modeling notation like EPC (event-driven process chain) or BPMN (Business Process Model and Notation). This chain of activities has to be executed, if all the prerequisites described in condition component are met.

### B. The CCC Model

Typically, modeling a new WPP is triggered by a workflow engineer or a privacy officer, who has identified a recurring, challenging problem which has no corresponding pattern. Recall that the WPP creator must be familiar with legal texts (Requirement R2), but the workflow engineer does not necessarily possess such knowledge. Thus, a model for deriving WPPs must ensure, that all legal obligations are included in the resulting WPP.

We will now outline the six steps needed to derive a WPP. They constitute our CCC Model. For this we use the structure described in previous subsection. We use Text 1 to illustrate

these steps. Note that Text 1 refers to an activity pattern. An example for a check pattern can be found in the Appendix.

a) *Define the Scope*: At first, the WPP creator sets the outline of the new WPP. He decides which legal articles and paragraphs will be in the scope. By setting the scope, he must ensure that the resulting WPP meets the requirements of design patterns (R1). In particular, the WPP must be not too complex or too simple to be useful. He also has to ensure that the new WPP can be combined with already existing WPPs (R3). Furthermore, the WPP creator has to consider that the legal texts in the scope do not contain uncertain legal concepts that are unsuitable for a WPP, as shown in Text 4. Scoping of a WPP can be supported with four questions:

- Is the scope suitable to create a WPP that is non-trivial?
- Is the scope understandable for the workflow engineer?
- Does the scope overlap with a WPP that already exists?
- Does the scope include legal texts that need to be interpreted individually by a legal expert?

*Example 1: The scope of the WPP is the implementation of the 'right of access' according GDPR for customer data. The company doesn't collect data from and doesn't transfer data to third parties, but it uses automated means for data processing of customer data in the EU. Furthermore, the WPP addresses only requests that arrive electronically.*

b) *Define the Header*: In this step, the meta-data of the WPP is defined. The meta-data of the pattern is the *Name*, the *Type*, the *Legal focus* and the *Relations to other WPPs*. The WPP name should be intuitively understandable and reflect the WPP type. A name beginning with 'Processing' would indicate an activity pattern, while a name starting with 'Lawfulness of' would refer to a check pattern.

The articles and paragraphs specified in 'Legal Focus' mirror the scope of the WPP. 'Relations to other WPPs' contains information if the scope of this WPP depends on, overlaps with or contradicts with existing WPPs.

*Example 2:*

**WPP Name** *Processing the Right of Access from the Inventory of Processing Activities*

**WPP Type** *Activity Pattern*

**Legal Focus** *Art. 15 Par. 1a-d, Par. 3; Art. 2 Par. 1; Art. 3 Par. 1 GDPR*

**Relation to other WPPs** *dependency to WPP 'Update Inventory of Processing Activities'*

c) *Define the Context*: In the third step, the context must be specified. It shall describe the situation and the purpose of the pattern in a plain language that is clearly understandable without legal expertise. It must provide answers for the following questions:

- Which business activities are in concern of this WPP?
- When does the privacy pattern apply?
- Which activities can occur before or after the WPP?

*Example 3: A business unit has received a request from a customer. The customer asks if personal data concerning him is processed. If this is the case, the customer must be given access to his personal data.*

d) *Define the Condition:* The condition translates legal requirements into prerequisites for the applicability of a WPP. The prerequisites have to be defined as positive statements that do not leave room for misunderstanding for a person without legal expertise. Thus, we discourage citing or referring to legal texts. The following questions serve as a guideline to obtain a check list of conditions:

- Which legal texts are in the 'Legal Focus' of this WPP?
- Do those texts base on other legal definitions?
- Which different requirements exist in each sentence of the legal text?
- Is a certain requirement already excluded by 'Context'?

*Example 4:*

- The identity of the requester has been verified.*
- The requester asks for his or her own data.*
- The requester does not make use of this right more than three times a year.*

e) *Define the Consequence:* The Consequence depends on the WPP type. For a check pattern only a state must be defined, which comes into effect when all requirements set in the Condition are met. For an activity pattern, the consequence is a chain of activities which must be specified (e.g. in form of an EPC notation) in this step.

*Example 5: Figure 4 describes the activities to process the request for access from a customer as a business process model.*

f) *Review the WPP:* To ensure that the pattern is correct and useful, it must be reviewed according to the following questions:

- Does the WPP meet the general quality criteria of design patterns?
- Is the WPP understandable and applicable for persons without legal expertise?
- Do the components Context, Condition and Consequence represent all information specified in the 'Legal Focus'?

### C. Discussion

We have derived our WPP representation from the general structure of legal texts. Essentially, we can represent any legal article (or its fragment) as a WPP. However, it was not in the scope of this paper to find out if a WPP representation makes sense for a certain use case. For example, Article 21 GDPR contains "legitimate grounds for the processing which override the interests, rights and freedoms of the data subject". It needs a lawyer to find out if such grounds indeed override the rights of the subject. If a WPP contains such concepts, it might not

be useful for a workflow engineer, who does not possess legal expertise. But it might be possible to decide upon such aspects at the creation time of the WPP. Thus, we see potential for further research.

It remains an open issue to evaluate our approach systematically. This is challenging: we have to consider three distinct user roles, with specific expertise areas and domain knowledge. It is difficult to separately assess the WPP representation and the framework for generating this representation. It is also challenging to exclude the properties of the application domain, when testing the applicability of a WPP. For this reason, we plan to evaluate our approach with a broad, qualitative case study.

Finally, it needs to be investigated how the creation, usage or verification of WPPs can be supported within workflow modeling tools or even within workflow modeling notations. Furthermore, corresponding frameworks and (semi-)automatic approaches would help to express the full potential of the WPPs. They could support the verification if the workflow embeds a WPP correctly. They also could help confirming if the WPP is conclusive, that is, if all (or particular) aspects of a certain legal text are represented within the WPP.

### V. CONCLUSION

The GDPR and other privacy norms resulted in new requirements for workflows that handle personal data. It may be - for example - a requirement to ensure that a particular information is used only for the purpose explained to the customer. This information must be deleted when the original purpose for which it was gathered is no longer valid. Furthermore, individual rights such as the 'right of access' or the 'right to be forgotten' require for new workflow extensions which are not directly related to the original core business objectives of a company.

Implementing privacy norms into workflows is challenging. Auditors, workflow engineers and data privacy officers normally have different fields of expertise, but must cooperate in an interdisciplinary way to implement or verify legal requirements in domain-specific business tasks. A promising approach to tackle such challenges is the use of Workflow Privacy Patterns (WPPs). WPPs provide solutions to problems recurring in enterprise workflows. However, existing work on WPPs does not explain how such patterns can be obtained in a systematic way.

In this paper, we have investigated how to derive WPPs from legal texts such as the GDPR. We have defined three distinct user roles that are involved in the creation and use of WPPs. Furthermore, we have compared the characteristics of legal texts with the properties of design patterns. From this point we have developed a formal representation of WPPs that follows the structure of legal norms. Furthermore, we have developed a framework that compiles WPPs in six steps. With two different use cases we have provided evidence that our approach allows to map articles of the GDPR into a formal representation which supports process engineers in designing workflows, which meet legal requirements.

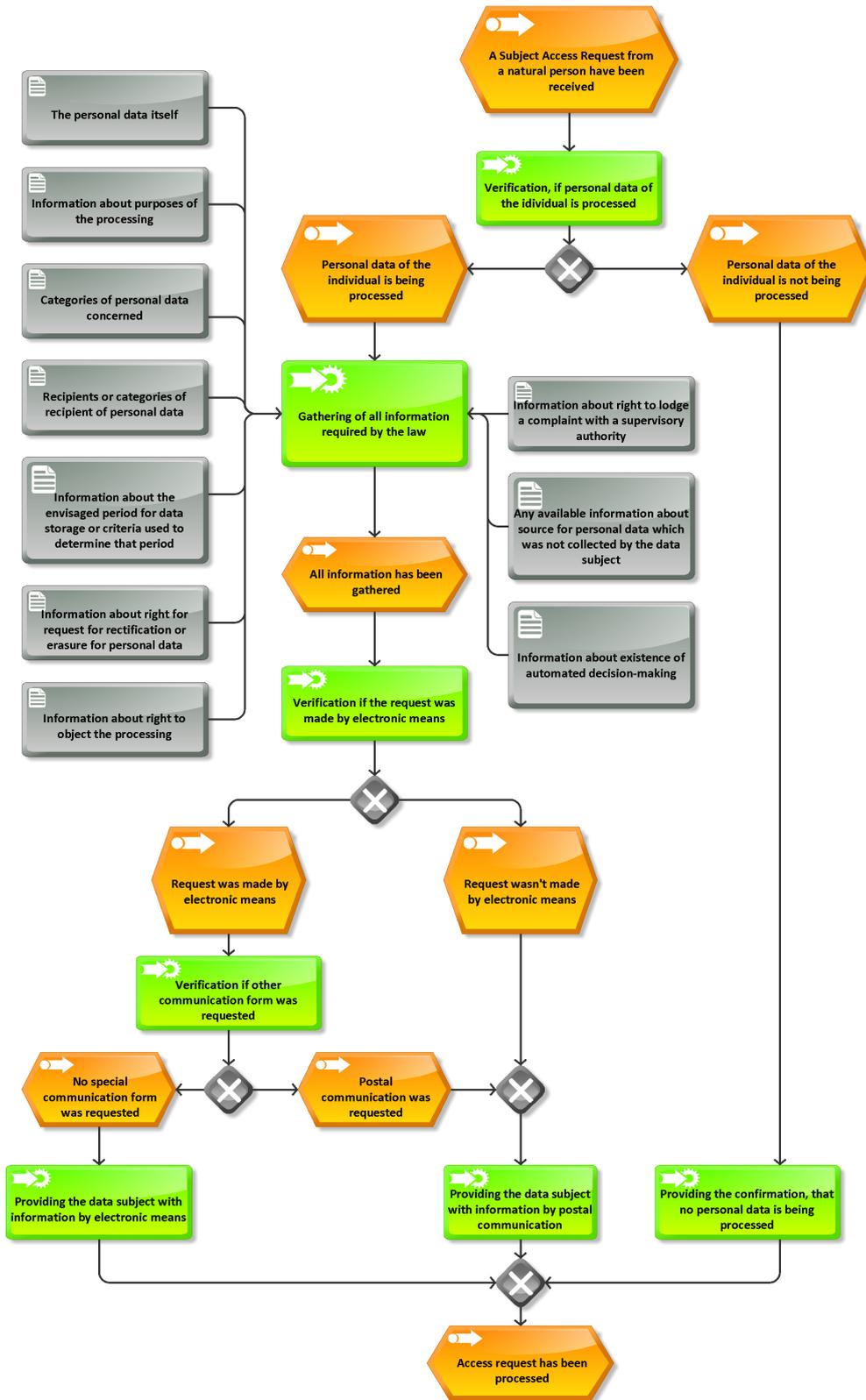


Fig. 4. Workflow to handle a request for access.

## ACKNOWLEDGMENT

We would like to thank Martin Bahr for his exceptional work on realizing the CCC Model.

APPENDIX  
EXAMPLE FOR A CHECK PATTERN

In this section, we illustrate a check pattern with the GDPR articles related to the consent for data processing. In particular, we have used our approach (cf. Section IV) to develop a WPP for the lawfulness of an electronic consent.

a) *Scope:*

Before an enterprise processes personal data, it must verify the lawfulness of processing. If there is no other legal basis, say, from other laws or a contract, the data subject must have been provided a consent to the processing of his or her data. The purpose of this WPP is to prove the lawfulness of an electronic consent from an adult according to the GDPR. The consent has been documented in a database.

b) *Header:*

**WPP Name** Lawfulness of an electronic consent  
**WPP Type** Check pattern  
**Legal Focus** The WPP considers the GDPR articles:

- Art. 4 ('definitions'), Par. 11 ('consent')
- Art. 6 ('lawfulness of processing'), Par. 1 (a)
- Art. 7 ('conditions for consent')

**Relation to other WPPs**

- 'Obtain Electronic Consent'
- 'Revoke Electronic Consent'

c) *Context:*

The purpose of this WPP is to prove the lawfulness of an electronic consent from an adult for the processing of personal data for a specific purpose.

d) *Condition:*

- There exists a record of a consent from the data subject in the database.
- The consent has been obtained in a lawful way. (cf. WPP 'Obtain Electronic Consent')
- The record documents that the data subject has been informed about processing activity, data to be processed, purpose of the processing, storage period, parties responsible for the processing and the receivers of the data.
- The record corresponds to the current processing.

- In the last 18 months, the consent has been given or there has been a processing activity related to this consent.
- There is an option to withdraw the consent that is easily accessible for the data subject. (cf. WPP 'Revoke Electronic Consent')
- The consent has not been withdrawn.

Note that the GDPR does not specify an expiration period for a consent. However, court decisions say that it is best practice not to rely on a consent that might have been forgotten already by the data subject.

e) *Consequence:*

If all conditions are fulfilled, a lawful consent for the processing exists.

## REFERENCES

- [1] European Parliament and Council of the European Union, "Regulation on the protection of natural persons with regard to the processing of personal data and on the free movement of such data," EU Regulation 2016/679, 2016.
- [2] E. Buchmann and J. Anke, "Privacy patterns in business processes," *INFORMATIK 2017*, 2017.
- [3] R. Von Alan and R. Hevner, "Design science in information systems research," *MIS quarterly*, 2004.
- [4] P. Schaar, "Privacy by design," *Identity in the Information Society*, vol. 3, no. 2, pp. 267–274, 2010.
- [5] Information Commissioners Office, "Guide to the general data protection regulation (gdpr)," <https://ico.org.uk>, Accessed Jul., 2018.
- [6] C. Alexander, *A pattern language: towns, buildings, construction*. Oxford university press, 1977.
- [7] P. Wolfgang, "Design patterns for object-oriented software development," *Reading Mass*, vol. 15, 1994.
- [8] D. C. Schmidt, M. Stal, H. Rohnert, and F. Buschmann, *Pattern-Oriented Software Architecture, Patterns for Concurrent and Networked Objects*. John Wiley & Sons, 2013, vol. 2.
- [9] A. Ter Hofstede, B. Kiepuszewski, A. Barros, and W. Aalst, "Workflow patterns," *Distributed and Parallel Databases*, vol. 14, no. 1, pp. 5–51, 2003.
- [10] S. Jablonski and C. Bussler, *Workflow management: modeling concepts, architecture and implementation*. International Thomson Computer Press London, 1996, vol. 392.
- [11] N. Russell, W. M. van der Aalst, and A. H. M. ter Hofstede, *Workflow Patterns: The Definitive Guide*. MIT Press, 2016.
- [12] N. Russell *et al.*, "Workflow control-flow patterns: A revised view," *BPM Center Report BPM-06-22*, *BPMcenter.org*, pp. 06–22, 2006.
- [13] —, "Workflow data patterns: Identification, representation and tool support," in *International Conference on Conceptual Modeling*. Springer, 2005, pp. 353–368.
- [14] —, "Workflow resource patterns: Identification, representation and tool support," in *International Conference on Advanced Information Systems Engineering*. Springer, 2005, pp. 216–232.
- [15] —, "Workflow exception patterns," in *Conference on Advanced Information Systems Engineering*, 2006.
- [16] B. S. Lerner *et al.*, "Exception handling patterns for process modeling," *Transactions on Software Engineering*, vol. 36, no. 2, 2010.
- [17] EU FP7 Project PRIPARE, "privacypatterns.eu - collecting patterns for better privacy," <https://privacypatterns.eu>, Accessed Apr., 2019.
- [18] Projects by IF, "Data permissions catalogue - an evolving collection of design patterns for sharing data," <https://catalogue.projectsbyif.com/>, Accessed Jun., 2019.
- [19] J. Vom Brocke, *Design principles for reference modeling: reusing information models by means of aggregation, specialisation, instantiation, and analogy*. IGI Global, 2007.
- [20] F. Buschmann, K. Henney, and D. C. Schmidt, *Pattern-oriented software architecture, on patterns and pattern languages*. John wiley & sons, 2007, vol. 5.



# Using Blockchain to Access Cloud Services: A Case of Financial Service Application

Min-Han Ruby Tseng

Graduate Institute of Technology Management  
National Chung Hsing University  
Taichung City, Taiwan  
Email:cv727320@gmail.com

Shuchih Ernest Chang\*

Graduate Institute of Technology Management  
National Chung Hsing University  
Taichung City, Taiwan  
Email:eschang@dragon.nchu.edu.tw

Tzu-Yin Kuo

Taipei Fubon  
Commercial Bank Co., Ltd.  
Taipei City, Taiwan  
Email:tzuyin9118@gmail.com

**Abstract**—Most cloud providers use centralized servers to manage data. However, centralized servers still suffer the risks of single point of failure and data theft. We add a blockchain to the cloud service and propose a new architecture to manage data. Using blockchain as a connector to utilize the tamperproof, traceable, and data-sharing features of the blockchain to ensure that the transaction data are properly stored in each node. We use the stock simulation trading service to extend and divide the research design into two levels, namely, system and application services. First, we directly write the data into the blockchain. Second, we alternatively store the data in the cloud and then write it into the blockchain. Finally, the two versions are compared and analyzed to investigate their feasibility and performance. At the application service, we implement the smart contract for the existing stock transaction process to achieve real-time settlement.

## I. INTRODUCTION

WITH the cloud service, the cost of equipment maintenance within the company is converted into the cost of service operation [1]. And the maintenance of the system becomes simple and also increases flexibility.

However, while availing of the convenience of cloud services, enterprises' internal data or even highly sensitive data are stored in the data center of a third-party. If sufficient security measures are not taken, then security risks, such as data leakage and tampering, will occur. In recent years, well-known cloud service providers have frequently reported cloud vulnerability incidents [2].

In particular, many small- and medium-sized enterprises (SMEs) adopt centralized third-party cloud services. As the size of the enterprise increases, the vertical expansion of the cloud service database becomes prone to the risk of single point of failure [3]. In the case of a single point of failure or single-path disconnection, the cloud service provider will interrupt the network service and even the entire production line. This situation can cause considerable losses for companies [4].

The centralized environment of enterprises is increasingly unable to adapt to the needs. Thus, they are gradually moving from the original centralized database to the decentralized database. The blockchain is a large decentralized database. The

data structure of the blockchain ensures that the transactions in the network are traceable, immutable, and tamperproof. In this study, the information stored in the cloud is encrypted, and the user's digital assets and transaction records are distributed in different nodes in the network through the P2P network, thereby reducing the risk of being stored only in a single node. Based on the blockchain, the data stored in the cloud is encrypted and stored in the network's block. The user's file data will not be exposed to the risk of being leaked or stolen during cloud server failure [5]. Users can securely access data, and privacy is well protected.

The design of this study will be divided into two levels, namely, system and application services. We classify the level discussed in the previous paragraph as the system service level and analyze the problems that cloud storage may encounter. We hope to reduce the risks that cloud storage may encounter by using the blockchain. At the application service level, we design the blockchain smart contract for the existing stock transaction process. The traditional stock transaction employs the T+2 settlement cycle. However, through the blockchain, users can achieve real-time delivery of stocks and the trading become more secure.

## II. RESEARCH BACKGROUND

This study is based on the cloud service and stock market transaction implementation services proposed by Wang and Chang [6]. We use the concept and technology of the blockchain to extend and improve the establishment of the blockchain as a connector for cloud data services.

### A. Stock market simulation trading system architecture

We design a network-based stock market simulation trading system (hereinafter called SMSTS). Using ASP.NET web development including HTML, CSS, Bootstrap, C# and Python to design the SMSTS. The database is designed using the Microsoft SQL Server.

### B. Blockchain

Satoshi Nakamoto published a paper entitled "Bitcoin: A Peer-to-Peer Electronic Funds Transfer System" in 2008 [7], proposing the concept of bitcoin and its underlying technology.

This work was supported by the Ministry of Science and Technology, Taiwan, under contract number MOST-106-2221-E-005-053-MY3.

\*Corresponding author: eschang@dragon.nchu.edu.tw

The blockchain is a large global decentralized ledger database that records all transaction records [8]. Each node uses the proof-of-work hash function to determine who verifies these transactions. The node that obtains the verification right would broadcast the block to all of the nodes. Until the first successful node confirms the verification, the block quickly connects to the parent blockchain.

### C. Smart contract

The concept of a smart contract was first proposed by Nick Szabo in 1994 [9]. He advocated that the trading conditions could be automated by the program. When the conditions are met, the value can be transferred. All of these are performed automatically by the computer program, and no third party is involved.

### D. Ethereum

The concept of Ethereum was first proposed by Vitalik Buterin: A next-generation smart contract and decentralized application [10]. Ethereum is an application platform based on blockchain technology that allow many different applications to be built on the Ethereum platform.

### E. Stock settlement

Settlement is the end of a creditor–debt relationship. If it cannot be completed in time, then it may cause the next transaction to be unsuccessful, which will affect other business activities. The stock settlement cycle is T+2. Thus, if the settlement speed can be improved, then the efficiency of the capital market operation can be improved and the cost of verification by the settlement institution personnel can be reduced.

## III. ARCHITECTURE AND DESIGN

In recent years, cloud computing has become even more popular. However, when confidential information is in third-party cloud services, the risk of leakage increases. The blockchain database consists of several nodes and all participate in data management. Any data added to the blockchain database must be agreed upon by most nodes in the blockchain network to be successfully recorded in the block and cannot be controlled by a single entity. Such mechanism ensures that data are secure, transparent, and permanently recorded, thus making it difficult to tamper with the content.

This study divides the blockchain into two levels, namely, system and application service levels. In terms of system services, two versions are proposed for writing data into the blockchain database. The first version involves directly writing the data into the blockchain. The second version involves storing the data in the cloud, and the blockchain acts as a connector to encrypt and decrypt the location where the cloud stores data. In terms of application services, the details of the use of smart contract automation for SMSTS are described in the subsequent paragraphs.

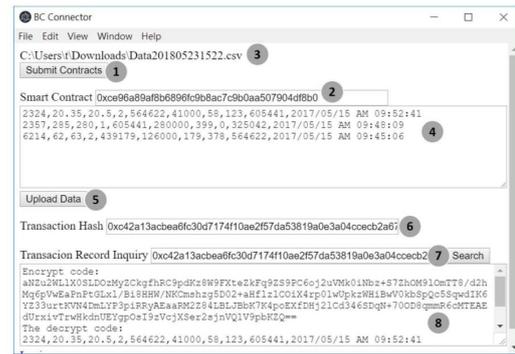


Fig. 1. User Interface of BC Connector

### A. Data directly stored in the blockchain

We use SMSTS that we have set up on the local server before, then set up the blockchain environment on the notebook. In this program, we designed the user interface to gradually write the data into the blockchain. Besides, we use symmetric key cryptography to encrypt the data.

1) *Blockchain environment setting*: We use the Ethereum platform to set up a private chain network and Solidity to write the smart contracts. Before the smart contract is submitted to the blockchain network, it needs to be compiled and deployed.

2) *Download the file*: We add the function button for downloading files in the SMSTS which enables users to download the transaction records from the SMSTS to their own computers and then write into the blockchain through the blockchain connector.

3) *Compile and deploy the smart contract*: To ensure easy operation by the user, this study designed a user interface that can be operated step by step, shown in Fig. 1. First, after pressing the (1) Submit Contracts button, it will generate an address in the (2) Smart Contract field, then click on File in the upper function bar and click on Open File, and then select the file to be written into the blockchain. After confirming the file, the file path will be displayed in (3). The file content will be displayed in the data column of (4). After pressing (5) Upload Data, the data will be written into the block. When the data are written into the blockchain, the encrypted Transaction Hash will be generated in (6). The user can also query the data of the blockchain through the (7) Transaction Record Inquiry, and the query data will be displayed in (8) Area. To ensure that the data are secure, we use the AES-128-CBC symmetric key encryption method to encrypt the data to be written into the blockchain. The Encrypt Code displayed in the query data bar is the encrypted garbled code. Even if the user obtains the Transaction Hash, the file content will still be invisible in the blockchain.

### B. Write the data address stored in the cloud into the blockchain

Most users use the centralized cloud server to store data, which is vulnerable to hackers; thus, the security of data is

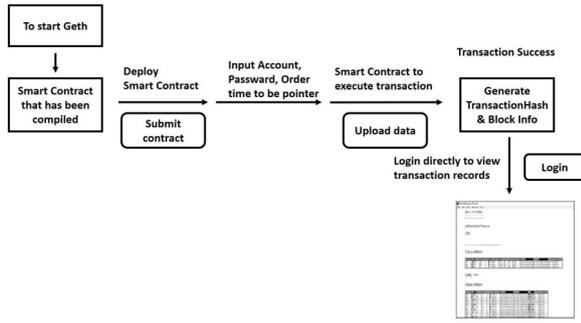


Fig. 2. Using the BC Connector to enter SMSTS query data

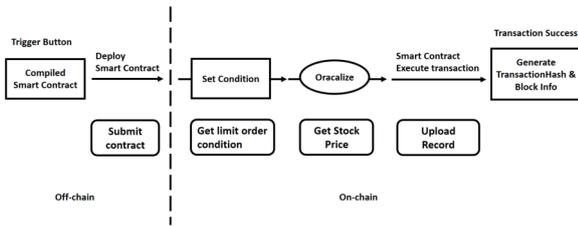


Fig. 3. Limit trading design flow

threatened. Therefore, this study integrates the web server, database, stock price program, and transaction processing program into the cloud system, allowing users to buy and sell stocks through the browser, thereby collecting data and using these data as the test data for writing the blockchain.

1) *Cloud erection*: We transfer the SMSTS to the Azure cloud platform and collect the data generated by the user through the simulation program for stock trading as the experimental basis.

2) *Using the blockchain as a connector*: We build a Electron, to design and develop our blockchain connector user interface. After launching Geth, another Git Bash window opens, with the electron command to execute our project. The account, password, and order time are used as pointers for the user on how to store the transaction data. After being encrypted and written into the blockchain, the user can directly log in to the transaction record page of the SMSTS to query the transaction data of the specified order date, as shown in Fig. 2.

C. Smart contract automation

Smart contract automation is applied to the stock limit trading service. Based on the design structure of the SMSTS, the system automatically generates the smart contract and deploys it to the P2P network environment. We have added a trigger button that allows users to link to their own blockchain wallet to perform limit trading services. The system design flow for this architecture is shown in Fig. 3

TABLE I  
APPLICATION LOAD TIME TEST RESULTS SET ON THE LOCAL AND CLOUD

Load time	SMSTS is set up on the local server (sec)	SMSTS is set up in Azure cloud server (sec)
1	2.07s	1.03s
2	2.54s	1.02s
3	2.05s	1.23s
4	2.01s	1.11s
5	2.52s	1.02s
6	2.40s	1.08s
7	2.31s	1.15s
8	2.23s	1.24s
9	1.99s	1.09s
10	2.36s	1.15s
Avg time	2.48s	1.12s

IV. EXPERIMENT AND EVALUATION

This study conducts a series of tests and analyses based on the architecture and design at the system service level described in the previous section, and the application service level will be compared with the existing platform.

A. Results from stock market simulation trading system test results

To test the performance, we use the Pingdom Website Speed Test, a free website tool that detects website speed and performance, to understand and evaluate the performance of the proposed service and to measure the load time of the website through experiments. We have designated San Jose(California,USA) as the test area.

The test results of the page load time and the time of writing data into the blockchain are shown in Table I. From the table, we can see that the cloud server is faster than the local server. This finding can be plausibly attributed to the fact that the cloud service providers focus more on optimization of the cloud storage service. This difference is also affected by the network speed, test location, and device capabilities between the cloud service provider and the research computer device.

B. Results from blockchain application system test results

We first compare the difference between (1)directly storing the accessing data in the cloud and (2)storing/accessing via cloud addresses recorded on blockchain. Table II shows that the time of transaction data generated by the SMSTS, which is directly stored in the cloud, is shorter and more average than that of other systems. Storing of the cloud data address to the blockchain increases the step of writing the blockchain; thus, it takes a long time. Compared with storing only the data in the cloud, writing into the blockchain is time consuming but more secure.

C. Discussion of cons and pros of the proposed blockchain system services

Traditional decentralized systems generally use a Replicated state machine [11] to implement a fault-tolerant mechanism. The blockchain uses similar approach, but it does not rely on a single entity to complete the service because there exists consensus agreement in the blockchain. When a transaction conflict occurs, only one transaction is approved to avoid double-spending. In order to solve the problem of Byzantine failure [12], we write the data encryption program in the smart

TABLE II  
COMPARISON OF DATA STORED IN THE CLOUD AND WRITE CLOUD DATA  
ADDRESS INTO THE BLOCKCHAIN

Load time	Data directly stored in the Azure cloud platform (sec)	Writes into the blockchain after the Azure output data address (sec)
1	21.14s	51.42s
2	20.31s	64.36s
3	15.92s	40.55s
4	14.54s	52.31s
5	24.64s	36.96s
6	18.51s	88.49s
7	16.61s	37.12s
8	19.92s	35.71s
9	23.16s	21.33s
10	17.18s	46.40s
Avg time	19.19s	47.45s

contract and enable it to be executed automatically, and then use the blockchain as a connector to achieve decentralized storage as the core of the cloud storage [13].

Through the solutions proposed in this study, the blockchain will be continuously extended, and the nodes can be connected to each other [14]. Once the data are written into the block, it cannot be tampered with, which helps the cloud provider in ensuring the security of the user data. Moreover, the nodes that are distributed in the network can reduce the cost of network transaction, authentication, and collaboration and can effectively solve the synchronization problem of the traditional distributed database. However, the current transaction speed of Ethereum is still very slow, only 10~20 transactions per second use sharding or plasma as a blockchain expansion solution.

#### D. Discussion of the benefit of blockchain application services

The proposed application is based on the limit trading service in the SMSTS developed in this research. Through the characteristics of the blockchain, real-time delivery of stocks can be achieved. The settlement cost can be reduced and the stock or payment time can be shortened [15].

#### V. CONCLUSION

This research is divided into three parts. First, we set up the web server and database system of the stock market simulation program in our Internet data center, collect transaction information from it, and store the data directly in our private blockchain. Moreover, the storage, verification, transmission, and communication of network data are performed through distributed nodes. Our private chain can record, sort, and encrypt every transaction. Participants use the verification code to link the transaction records and then use the characteristics of the blockchain to save records of all transactions and ensure the integrity of the data. Thus, the transaction history cannot be falsified.

Secondly, as more SMEs turn to cloud computing services, the blockchain can create secure, effective, tamperproof, and democratic computing networks. In this study, the location of each data block is recorded in the blockchain. When the file needs to be accessed, the system will verify the identity according to the private key of the user and assemble the file. We found that combining data in a blockchain with a decentralized cloud is safer than storing it in a centralized system. One device does not contain complete files, which

makes it almost impossible for hackers to steal data, thus improving security and reliability.

Third, the smart contract can automate of the stock limit trading service. The actual stock transactions do not need to wait for the T+2 settlement cycle. The blockchain increases the flexibility of trading strategy execution and reduces the labor cost of transaction clearance and settlement.

The blockchain can effectively reduce the cost of authentication, network transactions, and collaboration [16], can control the reading and modification of data through public and private keys by taking into account transparency and privacy security, and is reliable. This study conducts preliminary experiments in using the blockchain as a connector to store cloud data, and in the future, it can be applied to other applicable cloud services, even in different application scenarios.

#### REFERENCES

- [1] J. Gibson and R. Rondeau and D. Eveleigh and Q. Tan , "Benefits and challenges of three cloud computing service models," in *2012 Fourth International Conference on Computational Aspects of Social Networks (CASoN) IEEE*, 2012, pp. 198–205. DOI: 10.1109/CASoN.2012.6412402
- [2] H. Tianfield , "Security issues in cloud computing," in *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC) IEEE* 2012, pp. 1082–1089. DOI: 10.1109/ICSMC.2012.6377874
- [3] W.A. Jansen , "Cloud hooks: Security and privacy issues in cloud computing," in *2011 44th Hawaii International Conference on System Sciences IEEE* 2011, pp. 1–10. DOI: 10.1109/HICSS.2011.103
- [4] A. Kirar and A.K. Yadav and S. Maheswari, "An efficient architecture and algorithm to prevent data leakage in Cloud Computing using multi-tier security approach," in *2016 International Conference System Modeling & Advancement in Research Trends (SMART) IEEE* 2016, pp. 271–279. DOI: 10.1109/SYSMART.2016.7894534
- [5] M. Dai and S. Zhang and H. Wang and S. Jin, "A low storage room requirement framework for distributed ledger in blockchain," *IEEE Access*, vol. 6, 2018, pp. 22970–22975. DOI: 10.1109/ACCESS.2018.2814624
- [6] C.-W. Wang and S.E. Chang, "Cloud service in stock trading game: Service virtualization, integration and financial application," in *2016 Eighth International Conference on Ubiquitous and Future Networks (ICUFN) IEEE*, 2016, pp. 857–862. DOI: 10.1109/ICUFN.2016.7537158
- [7] S. Nakamoto, Bitcoin: A peer-to-peer electronic cash system, 2008, Retrieved Nov 20, 2018 from <https://bitcoin.org/bitcoin.pdf>
- [8] M. Swan, Blockchain: Blueprint for a new economy, 2015, " O'Reilly Media, Inc."
- [9] N. Szabo, Smart contracts, 1994, unpublished. Retrieved Dec 15, 2018 from [http://www.fon.hum.uva.nl/rob/Courses/InformationInSpeech/CDROM/Literature/LOTwinterschool2006/szabo.best.vwh.net/smart\\_contracts\\_2.html](http://www.fon.hum.uva.nl/rob/Courses/InformationInSpeech/CDROM/Literature/LOTwinterschool2006/szabo.best.vwh.net/smart_contracts_2.html)
- [10] V. Buterin, "A next-generation smart contract and decentralized application platform," white paper, 2014.
- [11] F. B. Schneider, "Implementing fault-tolerant services using the state machine approach: A tutorial," *ACM Computing Surveys (CSUR)*, vol. 22, no. 4, 1990, pp. 299–319. DOI: 10.1145/98163.98167
- [12] L. Lamport, R. Shostak, and M. Pease, "The Byzantine generals problem," *ACM Transactions on Programming Languages and Systems (TOPLAS)*, vol. 4, no. 3, 1982, pp. 382–401. DOI: 10.1145/357172.357176
- [13] X. Xu et al., "The blockchain as a software connector," in *2016 13th Working IEEE/IFIP Conference on Software Architecture (WICSA), IEEE* 2016, pp. 182–191. DOI: 10.1109/WICSA.2016.21
- [14] D. Drescher, "Blockchain Basics: A Non-technical Introduction in 25 Steps," 1st edn, Apress, Frankfurt am Main, 2017
- [15] K.-H. Huang, Application of blockchain in the field of securities trading: settlement delivery, 2017, Retrieved March 3, 2019 from <http://www1.cof.nkfust.edu.tw/ezfiles/12/1012/img/1938/729871942.pdf>
- [16] N. Herbaut and N. Negru, "A model for collaborative blockchain-based video delivery relying on advanced network services chains," *IEEE Communications Magazine*, vol. 55, no. 9, 2018, pp. 70–76. DOI: 10.1109/MCOM.2017.1700117

# Predicting Automotive Sales using Pre-Purchase Online Search Data

Philipp Wachter  
University of Hohenheim  
Schwerzstr. 35, 70599 Stuttgart,  
Germany  
Email: philipp.wachter@uni-  
hohenheim.de

Tobias Widmer  
University of Hohenheim  
Schwerzstr. 35, 70599 Stuttgart,  
Germany  
Email: tobias.widmer@uni-  
hohenheim.de

Achim Klein  
University of Hohenheim  
Schwerzstr. 35, 70599 Stuttgart,  
Germany  
Email: achim.klein@uni-  
hohenheim.de

**Abstract**—Sales forecasting is an essential element for implementing sustainable business strategies in the automotive industry. Accurate sales forecasts enhance the competitive edge of car manufacturers in the effort to optimize their production planning processes. We propose a forecasting technique that combines keyword-specific customer online search data with economic variables to predict monthly car sales. To isolate online search data related to pre-purchase information search, we follow a backward induction approach and identify those keywords that are frequently applied by search engine users. In a set of experiments using real-world sales data and Google Trends, we find that our keyword-specific forecasting technique reduces the out-of-sample error by 5% as compared to existing techniques without systematic keyword selection. We also find that our regression models outperform the benchmark model by an out-of-sample prediction accuracy of up to 27%.

## I. INTRODUCTION

IMPROVING the accuracy of sales forecasts is an important business challenge for optimizing production planning. As a decisive component of planning processes, sales forecasts form the basis of managerial decision-making. The automotive industry is characterized by a complex and uncertain business environment forcing car manufacturers to constantly improve their supply chain efficiency to stay competitive [1]. Hence, sales forecasts have become an integral component of supply chain processes. Because automotive manufacturers have implemented built-to-forecast vehicle production systems [2], accurate predictions are indispensable to ensure efficient production processes, optimize inventory levels, and improve the overall market performance [3]. Moreover, increasing product individualization places ever-higher demands on business information systems [4] and material requirements planning [5]. Inaccurate predictions can lead to inventory shortages, overstocking or unsatisfied customer demands [6].

Forecasting the future demand for durable consumer goods such as cars is challenging for three reasons. First, reliable forecasting models must integrate accurate representations of the customer buying behavior. Potential customers typically engage in online searches to determine what car to buy. Searching for pre-purchase information is regarded as an integral element of the consumer's buying behavior [7].

Extensive online research applies in particular to the purchasing process of cars. About 50% of the customers spend more than ten hours to identify the best matching vehicle for their requirements [8]. Ernst and Young report that customers devote more time for online research per-purchase of a car than for any other product [9]. Customers use different keywords and combinations of keywords to determine their choice. However, the extent to which these keyword-specific search results affect the sales performance of car manufacturers is still not known. Hence, understanding the online search behavior of customers is critical to improve forecast models. Second, fluctuating macroeconomic factors have a significant impact on automobile sales [10]. If lagged effects of economic factors are not considered in forecasting models, the forecast accuracy is further impaired. Third, in addition to the seasonal demand pattern for cars, external factors such as marketing campaigns further complicate the forecasting process.

Prior research on sales forecasting has focused on rather simple techniques that incorporate historic sales data and/or socioeconomic variables but pay little attention to information reflecting customer search behavior [10]–[12]. Subsequent approaches use customer online search data to predict car sales. Choi and Varian (2009) study a model that incorporates Google search data [13]. Their findings provide econometric evidence that using Google data can enhance the prediction accuracy of car sales. As a consequence, Google search data have become an important element of sales forecasting in this field of research [14]–[16]. Although these approaches predict car sales based on customer search data, they do not systematically select the most relevant keywords used by customers, which might lead to sales of new cars.

Against this backdrop, we propose a novel forecasting technique that combines keyword-specific customer search behavior from Google Trends with a set of economic variables for sales prediction in the automotive industry using a regression approach. To identify the most relevant keywords that customers use in Google prior to purchasing a new car, we use a backward induction approach. By using Google Ads, we identify the most relevant keywords that customers used

<sup>1</sup>This work has been partially supported by the Federal Ministry of Economic Affairs and Energy under grant ZF4541001ED8.

in Google search in the context of buying a new car. We include keywords related to new car purchases and exclude keywords associated to post-purchase and other queries unrelated with pre-purchase searches. Then, we obtain the Google Trends monthly time series of the most relevant keywords for new car sales. To validate our proposal, we use a unique dataset of car sales of a large car manufacturer from 2004 to 2019.

We find that our proposed forecasting technique improves the out-of-sample prediction accuracy by up to 5% as compared to models based on the same Google Trends search data without systematic keyword selection. Furthermore, we find that our forecasting models improve the out-of-sample accuracy by up to 27% compared to well-accepted autoregressive benchmark models.

The remainder of this paper is organized as follows. The next section discusses related literature on forecasting using online search data. In section 3, we present our proposed forecasting technique. In section 4, we report the experimental evaluation and discuss our findings. We provide our conclusion in section 5.

## II. RELATED WORK

Online search engines are frequently used as a starting point for the online research [17], [18]. With a market share of 88.5%, Google is by far the most frequently used online search service in the USA [19]. Due to the huge amount of daily search queries, Google represents a “Treasure House for web data mining” and previous research has focused on the predictive power of the search data [20]. Beside their popularity, search engines provide the benefit that the collected search data is less biased as compared to other user generated online data. In contrast to the use of social media platforms, online research is conducted in private and the personal activity is not revealed to others resulting in a less biased user behavior [21]. In recent years, several studies made use of Google data to improve forecasting as well as nowcasting accuracies. While forecasting is defined as the prediction of future events, nowcasting refers to the prediction of “the present, very near future and the very recent past” [22].

One of the first attempts to integrate search query data into a prediction model was made in the field of epidemiology [23]–[25]. Ginsberg et al. were able to predict the weekly influenza activity with a time lag of one day as they discovered a high correlation between influenza-related search queries and the percentage of daily physicians visits in which a patient had influenza-like symptoms. Further publications focus on the prediction of country-specific unemployment rates [13], [26]–[30], stock market movements and returns [31], [32], travel activities [33], and housing sales [13], [34]. During recent years, Google search data was employed in a wide range of different contexts, thus demonstrating the broad scope of possible application.

### A. Prediction of car sales

The use of Google search data for the prediction of car sales or car registrations has raised significant attention in the literature. Chamberlin (2010), Seebach et al. (2011), Du and Kamakura (2012), and Choi and Varian (2012) were the first who examined the predictive power of Google search data in the context of car sales [35], [14], [36], [33]. They conclude that Google data reflect changes in the volume of car sales and appears to be an appropriate data source for prediction models.

Carrière-Swallow and Labbé (2013) propose an online search data index to improve nowcasting models, predicting automotive sales in Chile [15]. Although they observe a relatively low Internet usage among the Chilean population, the integration of Google data improved both in-sample and out-of-sample nowcasts. In the former case, the whole data sample is used to fit the model and the forecasted observations are part of this sample (in-sample). As an attempt to mimic real data constraints, in the latter case, only a subset of the data sample is used to fit the model of which the forecasted observations are not part of (out-of-sample) [37].

Barreira et al. (2013) examine the eligibility of Google search data as a predictor for car sales in four European countries (i.e., France, Italy, Portugal, Spain) [30]. In contrast to previous work, they find only little evidence that Google data improves the accuracy of the prediction models for the included countries.

Taking cars as an example of high-involvement products, Geva et al. (2015) aim to improve the accuracy of an out-of-sample forecast by combining forum data in form of social media mentions/sentiments and search data [21]. They find a significant improvement of the prediction accuracy if both data sources are included in the model as compared to forum data only. Moreover, they observe a stronger improvement of the prediction accuracy for value than for premium brands.

Benthaus and Skodda (2015) pursue a similar approach by combining search data with Twitter sentiment data [16]. The results are in line with the findings by Geva et al. as a combination of the two data sources leads to an improved accuracy, both in-sample and out-of-sample.

The findings of Wijnhoven and Plant (2017), however, indicate that social media sentiments only have a minor predictive power as compared to Google search data or social mentions [38]. Consequently, Wijnhoven and Plant propose to only incorporate Google data and social mentions in a prediction model.

Fantazzini and Toktamysova (2015) investigate the out-of-sample accuracy of multivariate models using Google search data and economic variables to predict monthly sales of several car brands in Germany [3]. They find that Google data-based prediction models outperform competing models especially for forecast horizons longer than 12 months.

Nymand-Andersen and Pantelidis (2018) investigate the usefulness of Google search data with respect to nowcasting of car sales in the euro area [39]. They highlight the predictive

capabilities of online search data; however, they also underscore the need to further improve the data quality.

### B. Motivation of the search engine user

Although the use of online search data for forecasting purposes has grown considerably, the search motivation of customers has so far received little attention. The impact of the search motive on the predictive quality of a search query index can be highlighted by the following example. The search term “Honda” comprises multiple search purposes such as gathering product information before purchase, gathering product information after purchase, and gathering news about the brand or the product. To use the Google Trends data as a predictor for new car sales, the search query index should only reflect search queries related to a purchase intention (i.e., pre-purchase search). Extracting pre-purchase searches from aggregated data, however, remains a challenge. One approach is to use appropriate search categories (e.g., vehicle shopping) to exclude searches unrelated to a purchase. Graevenitz et al. (2016) argue that the underlying algorithms might be altered over time or the customer search behavior changes with regard to the keyword use [40]. Instead, they develop a model that links distinct search motives to the search and sales data to estimate the effect of pre-purchase queries on car sales. Hu et al. (2014) pursue another approach and try to isolate pre-purchase searches by excluding terms associated to post-purchase and other non-new-car-shopping-related searches (e.g., “parts”, “repair”) [41]. Most of the studies discussed above use rather simple keyword combinations, which only comprise the brand and/or the model name (e.g., “Honda + Civic”) depending on the level of aggregation.

## III. FORECASTING TECHNIQUE

Our proposed forecasting technique for new car sales using Google Trends search data for most relevant keywords is described as a five-step process depicted in Fig. 1. In a first step, relevant keywords and data are collected. To account for seasonality, the data is transformed to obtain deseasonalized time series. In a preliminary analysis, we detect the time lag of the Google Trends data and the economic variables with the car sales data. To identify the Google Trends data with the highest predictive power, we perform both an in-sample and an out-of-sample regression analysis. In the last step, we develop several multivariable regression models and determine the respective in-sample and out-of-sample performance.

### A. Google Trends tool

In 2006, Google launched the search analysis website Google Trends. The publicly available tool provides information about aggregated individual searches expressed in a search volume index. Hence, Google does not report the data in absolute numbers but provides the relative popularity of a search term. The index is calculated by dividing the data points of a query by the total volume of searches of the geography and time range considered [42]. The query shares

are normalized, such that 100 indicates the highest query share of the whole period. Since the search volume indices are proportionated to time and space, Google Trends allows to compare the relative popularity of a query across different geographic locations and time intervals.

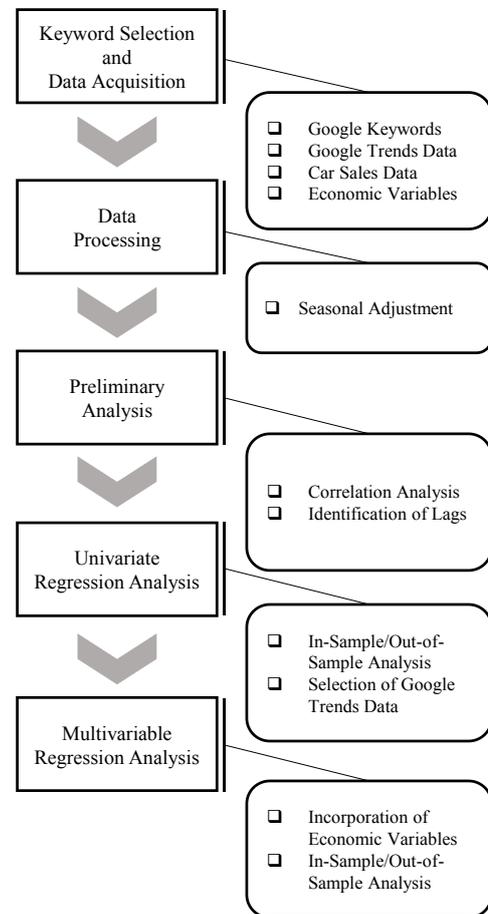


Fig. 1 Forecasting technique for car sales using most relevant search data

Moreover, Google introduced different categories and subcategories to refine the search for terms with multiple meanings. In the context of the automotive industry, the search results for “beetle” can be narrowed down by the choice of an appropriate category to exclude queries regarding the insect and only obtain results for the car offered by Volkswagen.

### B. Keyword selection

Selecting the most relevant keywords for Google Trends search is performed by using the online advertising platform Google Ads. Relevant keywords are identified following a backward induction approach [43]. The integrated Keyword planner tool suggests additional keywords based on keywords or groups of keywords entered by the user. The purpose of this process is to identify related keywords frequently employed by search engine users. We use the service to both identify top keywords that are commonly associated to new car purchase searches and keywords that relate to post-purchase or used car purchase searches.

### C. Data processing

Some data may exhibit a strong seasonality. To account for the systematic seasonal variation, we perform a decomposition operation. The time series is decomposed into a seasonally adjusted times series and the corresponding seasonal factors. This process is an implementation of the ratio-to-moving-average method (census method I). Due to the same reporting granularity the periodicity has not been adjusted.

### D. Preliminary analysis

This step encompasses a correlation analysis of the different Google Trends time series with the sales time series. As online information search is conducted in advance to new car shopping [44], we use cross-correlation to account for time lags. The incorporation of time lags is an essential prerequisite to obtain the optimal correlation between the data and to allow for forecasting the future instead of explaining the present. Cross-correlation has already been used to identify time lags in related previous work [14], [16]. The procedure is also applied to the selected economic variables. Moreover, the variables are checked for multicollinearity via bivariate Pearson correlation to prevent statistical and numerical issues in our subsequent regression analysis [45].

### E. Regression analysis

To determine the predictive power of Google Trends, search data (independent variable) and car sales (dependent variable) are used to estimate univariate linear regression models. We measure the in-sample and out-of-sample performance to identify the model with the best fit. Time lags detected during the preliminary analysis are taken into account for the model computation. We apply two performance criteria to evaluate the quality of the linear regression models. Both, Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are frequently employed for model evaluations [14], [16].

MAE measures the average magnitude of errors in a set of data regardless of the direction of the errors. As a linear score, all the individual differences are weighted the same. As shown in formula (1), the absolute difference of actual sales at time  $t$  ( $y_t$ ) and the predicted sales at time  $t$  ( $\hat{y}_t$ ) is divided by the number of observations ( $n$ ).

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| \quad (1)$$

For the second regression metric, the error is also calculated as an average of the absolute differences between actual sales and predicted values, however, the individual deviations have been squared before. This leads to the fact that the RSME (see formula 2) is more sensitive to outliers.

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (2)$$

After identifying the Google Trends data with the best in-sample accuracy, we estimate additional multivariable linear regression models by including different combinations of economic variables.

As a benchmark, we use a seasonal autoregressive baseline model (see formula 3) previously applied in several studies [33], [39]. The model uses 12 months ( $S_{t-12}$ ) and 1-month ( $S_{t-1}$ ) lagged historic sales data and an error term  $\varepsilon_t$  to predict car sales  $S_t$ .

$$S_t = \beta_0 + \beta_1 S_{t-12} + \beta_2 S_{t-1} + \varepsilon_t \quad (3)$$

## IV. EVALUATION

This section reports an experimental evaluation of our forecasting technique for car sales based on most relevant Google Trends data. We describe the setup, report the results, and discuss the findings

### A. Experimental setup

We collected monthly search query indices for the respective car model and/or car brand in combination with the most relevant keywords selected via Google Ads. We focus on the car manufacturer Honda as a representative of a large seller in the US. To obtain Google Trends data for the brand Honda, we additionally include the model names of the four best-selling car models responsible for approximately 90% of the Honda car sales in the period considered. The intention is to achieve a high coverage of search queries for Honda cars by using the top sellers as a proxy. To exclude searches unrelated to the automotive industry the search query indices are generated within the category “Autos & Vehicles”. The result data are limited to searches originating from the US in the period from January 2004 to February 2019.

Our evaluation is based on a unique dataset containing the monthly US car sales from January 2004 to January 2016. We obtained additional data from February 2015 to February 2019 from the automotive industry analysis website CarSaleBase [46], which has been used as a source for automotive sales information in prior research [47]. To ensure the consistency of the two datasets, we check that the car sales data are congruent in 2015. We obtained 182 observations for each Honda car model.

The economic variables have been systematically selected on the basis of relevant literature [3], [48]–[50]. The variables either reflect changes in the price paid by automobile consumers, affect the automobile sales demand, or describe the state of the US economy [50]. Table I shows the selected variables and the respective descriptions.

TABLE I.  
ECONOMIC VARIABLES

Economic variable	Source	Description
Consumer confidence index (CCI)	OECD	The index provides a measure for the consumer confidence and indicates future developments regarding consumption and saving
Consumer price index for new vehicles (CPI)	BLS	The index reflects changes in the price level for new vehicles (base period 1982-1984=100)
Gasoline price	EIA	The monthly retail price of US regular all formulations gasoline price
Unemployment rate	BLS	US national unemployment rate
Standard & Poor's 500 Index (S&P 500)	Yahoo finance	US stock market benchmark

BLS: Bureau of Labor Statistics; EIA: Energy Information Administration; OECD: Organization for Economic Cooperation and Development

We used data from January 2004 to August 2017 to estimate the linear regression models. Consequently, this is also the period for the in-sample analysis. To evaluate the out-of-sample performance, we used the in-sample estimated models to predict car sales from September 2017 to February 2019.

As a linear relationship between the independent and the dependent variables is a fundamental prerequisite for a linear regression analysis, we verify linearity by using scatterplots. To ensure that the remaining assumptions are fulfilled, we analyzed the histogram of residuals and the P-P plot. To ensure homogeneity of variances, we examined a scatterplot of the predicted values and the residuals.

B. Results

We identified frequently employed keywords for searches relating to new car purchases and for searches not related to pre-purchase situations using the keyword planner tool of Google Ads. While pre-purchase keywords often relate to the procurement processes (e.g., search for car dealers), pre-purchase unrelated keywords predominantly cover attributes associated to used cars or car maintenance and repairs. Table II shows the different brand-related keyword sets, additional pre-purchase keywords, and the pre-purchase unrelated keywords that can be used for reducing search data results.

TABLE II.  
KEYWORDS USED FOR RETRIEVING GOOGLE TRENDS SEARCH DATA

	Brand-related keyword sets	Pre-purchase keywords	Pre-purchase unrelated keywords
1	honda	new + buy +	repair -tires -
2	civic + accord + crv + odyssey	dealers + dealerships + compare	mechanic - maintenance - inspection -old - used -owned - parts -lease
3	honda + civic + accord + crv + odyssey		

Table III shows the in-sample performance of the different regression models. We conducted an in-sample cross-correlation analysis to detect the optimal time lag between the Google Trends data and the sales data. For each Google Trends time series, the highest correlation was identified without any time lag. Our results indicate a positive relationship between Google Trends search data and car sales for all univariate linear regression models. The correlation coefficients ranged from 0.69 to 0.83 and were significant at  $p < 0.01$ . Search queries based on keywords for car models (set 2 in Table II) resulted in Google Trends data with the highest explanatory power in the in-sample analysis. The results also imply that specifying pre-purchase unrelated keywords to be excluded from search data further improves the model

TABLE III.  
IN-SAMPLE PERFORMANCE OF GOOGLE TRENDS BASED LINEAR REGRESSION MODELS

Keyword set (brand)	Keywords (pre-purchase)	Keywords (pre-purchase unrelated)	Correlation coefficient	Root mean squared error	Mean absolute error
1			0.70**	14845.5	11830.2
	●		0.69**	15133.2	11591.5
		●	0.72**	14603.9	11751.8
	●	●	0.69**	15150.8	11631.4
2			0.82**	11873.5	8864.5
	●		0.71**	14787.2	11687.4
		●	0.83**	11815.0	8856.0
3			0.71**	14819.7	11829.4
	●		0.69**	15228.5	11796.4
		●	0.79**	12952.9	10335.0

\*\* $p < 0.01$

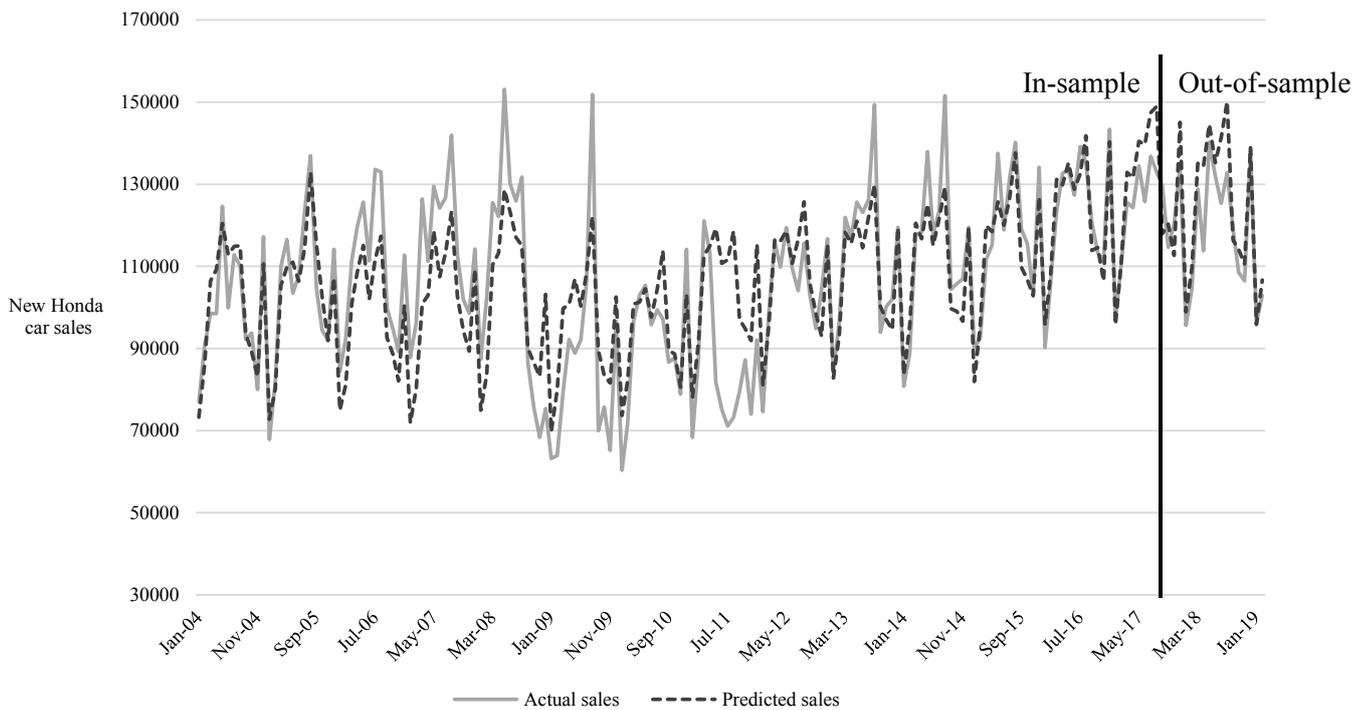


Fig. 2 Actual car sales and predicted sales based on most predictive Google Trends data

performance. That is, using car model keywords and specifying pre-purchase unrelated keywords for exclusion leads to lowest error measures among all regression models in the in-sample analysis (RMSE=11815; MAE=8856).

Excluding pre-purchase unrelated keywords from Google Trends data on car model keywords (set 2 in Table II) reduced the out-of-sample MAE by 5% from 7564.2 to 7183.8. Compared to Google Trends data on brand name (set 1 in Table II) without considering further keywords, the out-of-sample error (MAE=16796.8) is reduced by more than half. However, including keywords related to new car purchases do not reduce the prediction error. Fig. 2 illustrates actual sales and predicted sales using Google Trends data with the highest in-sample and out-of-sample accuracy. The figure demonstrates face validity of our approach.

After selecting the Google Trends data with the lowest prediction error, we conducted an out-of-sample analysis with a time horizon of 18 months. We included a set of economic variables to test for further reducing the prediction error. Since most of the economic variables are known to be leading or lagging indicators, we first identified the most predictive time lags via cross-correlation with the car sales data. Time lags were restricted to -12 to 0 months. If positive time lags for the variables were detected (i.e., economic variable from January 2016 has the highest correlation with car sales from December 2015), we incorporated no time lag. Table IV shows the chosen time lags for the economic variables and the corresponding correlation matrix.

TABLE IV.  
CORRELATIONS BETWEEN CAR SALES AND ECONOMIC VARIABLES

Variable	Optimal time lag in months	Sales	CPI	CCI	S&P 500	Unempl.
Sales						
CPI	-12	0.60**				
CCI	-10	0.57**	0.40**			
S&P 500	0	0.67**	0.91**	0.44**		
Unempl.	0	-0.63**	-0.38**	-0.89**	-0.50**	
Gasol. p.	0	0.04	0.69	-0.44**	0.04	0.39**

\*\*p<0.01; Unempl.: Unemployment; Gasol. p.: Gasoline price

All economic variables except the gasoline price showed a statistically significant correlation with car sales at  $p<0.01$ . The strongest correlation with car sales was observed for S&P 500 without time lag. Based on this preliminary analysis, we systematically generated univariate and multivariable linear regression models. The combination of predictors was restricted by the prevention of multicollinearity effects. Multicollinearity refers to a state of very high intercorrelation among the independent variables, which potentially impairs the unbiased estimation of the regression coefficients.

Because the gasoline price did not correlate with car sales in our analysis, the variable was not incorporated in any model. In addition to combinations of Google Trends data and 1-month lagged Google Trends data with each economic variable, we included all eligible combinations of economic variables.

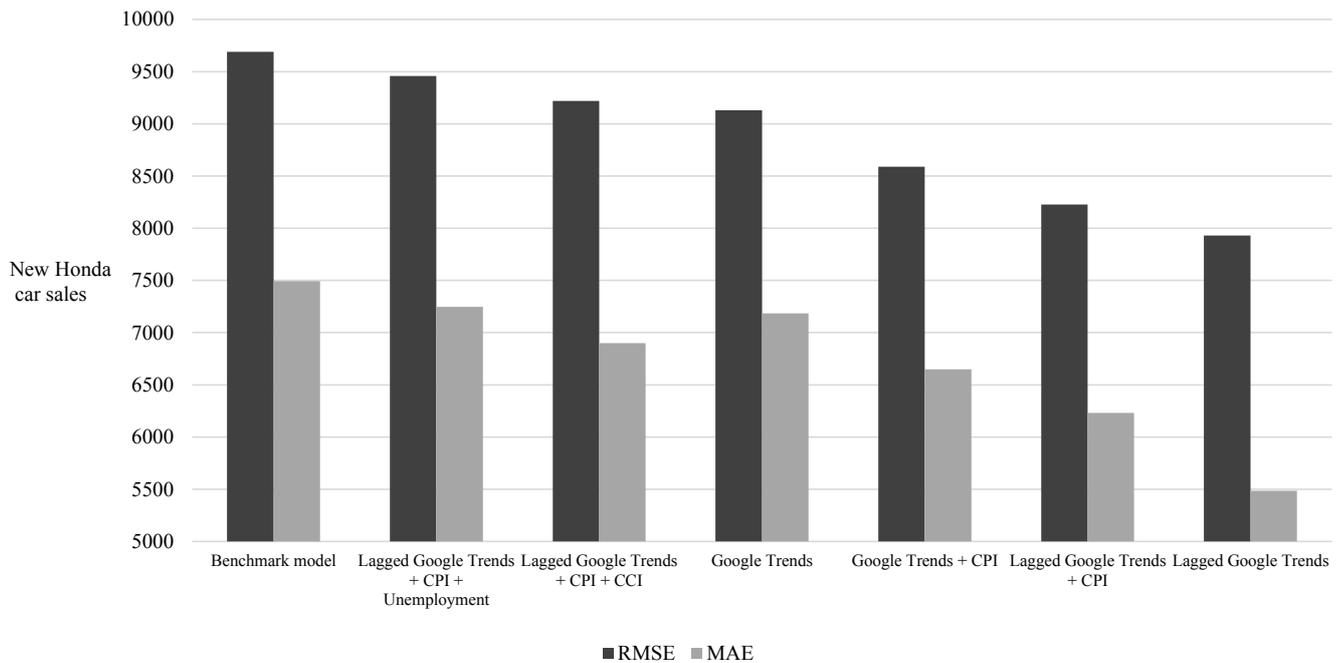


Fig. 3 Out-of-sample car sales prediction error of linear regression models with different predictive variables

Fig. 3 shows the results of the out-of-sample evaluation on a prediction horizon of 18 months. The figure only includes prediction models that outperformed the benchmark model. Both univariate Google Trends models achieved higher prediction accuracies compared to the benchmark model. A combination of Google Trends with economic variables did not necessarily improve the out-of-sample performance. A 12-months lagged CPI appeared to be the only predictor that decreases the forecasting error for unlagged Google Trends. All other multivariable regression models failed to improve the performance of the respective univariate Google Trends model. Although the in-sample error of Google Trends data with a time lag of one month (MAE=9536.5) is higher than that of Google Trends without time lag (MAE=8811.8), lagged Google Trends data achieved the smallest out-of-sample prediction error.

All (multiple) linear regressions met the criteria for linear regressions. For some of the linear regressions, we observed a slight bunching of the residuals, resulting in not perfectly identically distributed values. However, we considered the homoscedasticity assumption as fulfilled.

### C. Discussion

Our experiments demonstrate the effectiveness of carefully selected customer online search data from Google for accurately predicting automotive sales. Our findings provide evidence that our proposed forecasting technique benefits from the predictive power of Google Trends data. In the following paragraphs, we discuss the insights that can be obtained from our research.

Although most customers engage in online information search prior to the purchase of a new car, our results imply that Google Trends search data without any time lag yield the highest correlation with car sales. That is, Google Trends data

is most effective for predicting car sales of the current month. This finding is consistent with the results of prior research that identifies only few to no month(s) time lag [15], [16].

In the 18 months out-of-sample analysis, however, we find the highest performance using Google Trends data with a time lag of one month. Prior research suggests that, high in-sample prediction accuracy does not necessarily lead to high accuracy in an out-of-sample analysis and vice versa [14].

Our technique achieved the highest in-sample and out-of-sample accuracy for Google Trends data based on car model names combined with an exclusion of search queries containing keywords unrelated to pre-purchase situations. This finding becomes particularly evident in the out-of-sample analysis. Here, the prediction error was reduced by approximately 5% as compared to Google Trends data without keyword exclusion. Although adding pre-purchase associated keywords did not improve the model performance, systematic keyword use improved the predictive power of the Google Trends data in general.

While we find, with one exception, that incorporating the selected economic variables does not reduce the out-of-sample error, the in-sample performance was generally improved by adding the economic variables. While for the basic Google Trends data, combinations with both CCI and unemployment rate reduce the in-sample error, any two-variable combination of Google Trends data with a time lag of one month with one of the economic variables (CPI, CCI, S&P 500, unemployment rate) improves the in-sample performance. Moreover, any tested three-variable combination (Google Trends + economic variable 1 + economic variable 2) outperformed the respective univariate Google Trends regression model in the in-sample analysis. As depicted in Fig. 3, several multivariable regression models attained smaller prediction errors in the out-of-sample

analysis as compared to the benchmark model. A combination of Google Trends data with economic variables, however, did not always improve the accuracy of the corresponding univariate Google Trends model.

Future research might be pursued in at least two directions. First, while we focus on top keywords proposed by the keyword planner tool of Google Ads in this work, integrating additional keywords and keyword combinations could further improve the accuracy of the prediction. These additional keywords could be obtained by empirical studies that focus on customer search behavior. Second, although our experimental setup appears to be sufficient for our research purpose, more sophisticated methods for sales forecasting are available. Hence, our approach might be extended to machine learning methods such as Neural Networks.

## V. CONCLUSION

Our findings imply that predictions based on most relevant Google Trends search data that exclude pre-purchase unrelated searches improve the out-of-sample accuracy by up to 5% as compared to Google Trends data without systematic keyword selection. Moreover, we combine Google Trends data with relevant economic variables commonly employed for new car sales forecasting. In the performance evaluation of our linear regression models against a common seasonal autoregressive benchmark model, we find an improvement of the out-of-sample accuracy of up to 27%. Our findings help car manufacturers to obtain better forecasts and to make more informed decisions concerning their business strategies for production planning.

## ACKNOWLEDGMENT

We thank Hansjörg Tutsch and Joerg Leukel for their valuable comments on earlier versions of this paper.

## REFERENCES

- [1] J.-H. Thun and D. Hoenig, "An empirical analysis of supply chain risk management in the German automotive industry," *International Journal of Production Economics*, vol. 131, no. 1, pp. 242–249, 2011, <http://dx.doi.org/10.1016/j.ijpe.2009.10.010>.
- [2] J. Roehrich, G. Parry, and A. Graves, "Implementing build-to-order strategies: enablers and barriers in the European automotive industry," *International Journal of Automotive Technology and Management*, vol. 11, no. 3, pp. 221–235, 2011, <http://dx.doi.org/10.1504/IJATM.2011.040869>.
- [3] D. Fantazzini and Z. Toktamysova, "Forecasting German car sales using Google data and multivariate models," *International Journal of Production Economics*, vol. 170, pp. 97–135, 2015, <http://dx.doi.org/10.1016/j.ijpe.2015.09.010>.
- [4] J. Leukel, A. Jacob, P. Karaenke, S. Kirn, and A. Klein, "Individualization of goods and services: towards a logistics knowledge infrastructure for agile supply chains," in *Proceedings of the 2011 AAAI Spring Symposium on AI for Business Agility*, Stanford, CA, USA, 2011, pp. 36–49.
- [5] T. Widmer, A. Klein, P. Wachter, and S. Meyl, "Predicting Material Requirements in the Automotive Industry Using Data Mining," in *Business Information Systems*, Seville, Spain, 2019, pp. 147–161.
- [6] G. Nunnari and V. Nunnari, "Forecasting Monthly Sales Retail Time Series: A Case Study," in *2017 IEEE 19th Conference on Business Informatics (CBI)*, Thessaloniki, Greece, 2017, pp. 1–6.
- [7] K. Akalamkam and J. K. Mitra, "Consumer Pre-purchase Search in Online Shopping: Role of Offline and Online Information Sources," *Business Perspectives and Research*, vol. 6, no. 1, pp. 42–60, 2018, <http://dx.doi.org/10.1177/2278533717730448>.
- [8] K. Kandaswam and A. Tiwar. "Driving through the consumer's mind: Steps in the buying process." <https://www2.deloitte.com/content/dam/Deloitte/in/Documents/manufacturing/in-mfg-dtcm-steps-in-the-buying-process-noexp.pdf> (accessed Apr. 18, 2019).
- [9] EY. "Future of automotive retail Shifting from transactional to customer-centric." <https://www.ey.com/Publication/vwLUAssets/EY-future-of-automotive-retail/%24FILE/EY-future-of-automotive-retail.pdf> (accessed Apr. 18, 2019).
- [10] S. Shahabuddin, "Forecasting automobile sales," *Management Research News*, vol. 32, no. 7, pp. 670–682, 2009, <http://dx.doi.org/10.1108/01409170910965260>.
- [11] R.M.J. Heuts and J.H.J.M. Bronckers, "Forecasting the Dutch heavy truck market," *International Journal of Forecasting*, vol. 4, no. 1, pp. 57–79, 1988, [http://dx.doi.org/10.1016/0169-2070\(88\)90010-6](http://dx.doi.org/10.1016/0169-2070(88)90010-6).
- [12] F.-K. Wang, K.-K. Chang, and C.-W. Tzeng, "Using adaptive network-based fuzzy inference system to forecast automobile sales," *Expert Systems with Applications*, vol. 38, no. 8, pp. 10587–10593, 2011, <http://dx.doi.org/10.1016/j.eswa.2011.02.100>.
- [13] H. Choi and H. Varian, "Predicting the Present with Google Trends," *Google Inc*, 2009.
- [14] C. Seebach, I. Pahlke, and R. Beck, "Tracking the Digital Footprints of Customers: How Firms can Improve their Sensing Abilities to Achieve Business Agility," *Proceedings of the 19th European Conference on Information Systems (ecis)*, 2011.
- [15] Y. Carrière-Swallow and F. Labbé, "Nowcasting with Google Trends in an Emerging Market," *Journal of Forecasting*, vol. 32, no. 4, pp. 289–298, 2013, <http://dx.doi.org/10.1002/for.1252>.
- [16] J. Benthous and C. Skodda, "Investigating consumer information search behavior and consumer emotions to improve sales forecasting," in *Proceedings of the 21st Americas Conference on Information Systems*, Puerto Rico, 2015.
- [17] J. Otterbacher, "Searching for product experience attributes in online information sources," in *Proceedings of the International Conference on Information Systems (ICIS 2008)*, 2008, paper 207.
- [18] N. Kumar, K. R. Lang, and Q. Peng, "Consumer Search Behavior in Online Shopping Environments," in *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, Big Island, HI, USA, Jan. 2005, 175b–175b.
- [19] statcounter. "Search Engine Market Share United States Of America." <http://gs.statcounter.com/search-engine-market-share/all/united-states-of-america> (accessed Apr. 18, 2019).
- [20] L. Vaughan and Y. Chen, "Data mining from web search queries: A comparison of google trends and baidu index," *Journal of the Association for Information Science and Technology*, vol. 66, no. 1, pp. 13–22, 2015, <http://dx.doi.org/10.1002/asi.23201>.
- [21] T. Geva, G. Oestreich-Singer, N. Efron, and Y. Shimshoni, "Using forum and search data for sales prediction of high-involvement products," *MIS Quarterly*, vol. 41, no. 1, pp. 65–82, 2017, <http://dx.doi.org/10.25300/MISQ/2017/41.1.04>.
- [22] M. Banbura, D. Giannone, and L. Reichlin, "Nowcasting," *ECB Working Paper No. 1275*, 2010.
- [23] J. Ginsberg *et al.*, "Detecting influenza epidemics using search engine query data," *Nature*, vol. 457, no. 7232, pp. 1012–1014, 2009, <http://dx.doi.org/10.1038/nature07634>.
- [24] P. M. Polgreen, Y. Chen, D. M. Pennock, and F. D. Nelson, "Using internet searches for influenza surveillance," (eng), *Clinical Infectious Diseases : an official publication of the Infectious Diseases Society of America*, vol. 47, no. 11, pp. 1443–1448, 2008, <http://dx.doi.org/10.1086/593098>.
- [25] A. F. Dugas *et al.*, "Influenza forecasting with Google Flu Trends," (eng), *PLoS one*, vol. 8, no. 2, e56176, 2013, <http://dx.doi.org/10.1371/journal.pone.0056176>.
- [26] J. Pavlicek and L. Kristoufek, "Nowcasting unemployment rates with google searches: Evidence from the visegrad group countries," *PLoS one*, vol. 10, no. 5, e0127084, 2015, <http://dx.doi.org/10.1371/journal.pone.0127084>.
- [27] Y. Fondeur and F. Karamé, "Can Google data help predict French youth unemployment?," *Economic Modelling*, vol. 30, no. C, pp. 117–125, 2013, <http://dx.doi.org/10.1016/j.econmod.2012.07.017>.

- [28] N. Askitas and K. F. Zimmermann, "Google econometrics and unemployment forecasting," *Applied Economics Quarterly*, vol. 55, no. 2, pp. 107–120, 2009, <http://dx.doi.org/10.2139/ssrn.1465341>.
- [29] F. D'Amuri and J. Marcucci, "The predictive power of Google searches in forecasting US unemployment," *International Journal of Forecasting*, vol. 33, no. 4, pp. 801–816, 2017, <http://dx.doi.org/10.1016/j.ijforecast.2017.03.004>.
- [30] N. Barreira, P. Godinho, and P. Melo, "Nowcasting unemployment rate and new car sales in south-western Europe with Google Trends," *NETNOMICS: Economic Research and Electronic Networking*, vol. 14, no. 3, pp. 129–165, 2013, <http://dx.doi.org/10.1007/s11066-013-9082-8>.
- [31] T. Preis, H. S. Moat, and H. E. Stanley, "Quantifying trading behavior in financial markets using Google Trends," *Scientific reports*, vol. 3, p. 1684, 2013, <http://dx.doi.org/10.1038/srep01684>.
- [32] L. Bijl, G. Kringhaug, P. Molnár, and E. Sandvik, "Google searches and stock returns," *International Review of Financial Analysis*, vol. 45, no. C, pp. 150–156, 2016, <http://dx.doi.org/10.1016/j.irfa.2016.03.015>.
- [33] H. Choi and H. Varian, "Predicting the Present with Google Trends," *Economic Record*, vol. 88, no. 1, pp. 2–9, 2012, <http://dx.doi.org/10.1111/j.1475-4932.2012.00809.x>.
- [34] L. Wu and E. Brynjolfsson, "Chapter 3 - The Future of Prediction," in *Economic Analysis of the Digital Economy*, A. Goldfarb, S. M. Greenstein, and C. E. Tucker, Eds.: University of Chicago Press, 2015, pp. 89–118.
- [35] G. Chamberlin, "Googling the present," *Economic & Labour Market Review*, vol. 4, no. 12, pp. 59–95, 2010, <http://dx.doi.org/10.1057/elmr.2010.166>.
- [36] R. Y. Du and W. A. Kamakura, "Quantitative trendspotting," *Journal of Marketing Research*, vol. 49, no. 4, pp. 514–536, 2012, <http://dx.doi.org/10.1509/jmr.10.0167>.
- [37] A. Inoue and L. Kilian, "In-Sample or Out-of-Sample Tests of Predictability: Which One Should We Use?," *Econometric Reviews*, vol. 23, no. 4, pp. 371–402, 2005, <http://dx.doi.org/10.1081/ETC-200040785>.
- [38] F. Wijnhoven and O. Plant, "Sentiment analysis and Google trends data for predicting car sales," in *38th International Conference on Information Systems*, 2017.
- [39] P. Nyman-Andersen and E. Pantelidis, "Google econometrics: nowcasting euro area car sales and big data quality requirements," ECB Statistics Paper, 2018.
- [40] G. von Graevenitz, C. Helmers, V. Millot, and O. Turnbull, "Does Online Search Predict Sales? Evidence from Big Data for Car Markets in Germany and the UK," *CGR Working Paper*, 2016, <http://dx.doi.org/10.2139/ssrn.2832004>.
- [41] Y. Hu, R. Y. Du, and S. Damangir, "Decomposing the Impact of Advertising: Augmenting Sales with Online Search Data," *Journal of Marketing Research*, vol. 51, no. 3, pp. 300–319, 2014, <http://dx.doi.org/10.1509/jmr.12.0215>.
- [42] Google. "How Trends data is adjusted." [https://support.google.com/trends/answer/4365533?hl=en&ref\\_topic=6248052](https://support.google.com/trends/answer/4365533?hl=en&ref_topic=6248052) (accessed Apr. 18, 2019).
- [43] A. Ross, "Nowcasting with Google Trends: a keyword selection method," *Fraser of Allander Economic Commentary*, vol. 37, no. 2, pp. 54–64, 2013.
- [44] L. R. Klein and G. T. Ford, "Consumer search for information in the digital age: An empirical study of prepurchase search for automobiles," *Journal of Interactive Marketing*, vol. 17, no. 3, pp. 29–49, 2003, <http://dx.doi.org/10.1002/dir.10058>.
- [45] A. F. Siegel, "Multiple Regression," in *Practical Business Statistics*: Elsevier, 2016, pp. 355–418.
- [46] Carsalebase. "Automotive Industry analysis, opinions and data." <carsalebase.com/> (accessed Apr. 18, 2019).
- [47] K. Afrin, B. Nepal, and L. Monplaisir, "A data-driven framework to new product demand prediction: Integrating product differentiation and transfer learning approach," *Expert Systems with Applications*, vol. 108, pp. 246–257, 2018, <http://dx.doi.org/10.1016/j.eswa.2018.04.032>.
- [48] M. Hülsmann, D. Borscheid, C. M. Friedrich, and D. Reith, "General Sales Forecast Models for Automobile Markets and their Analysis," *Transactions on Machine Learning and Data Mining*, vol. 5, no. 2, pp. 65–86, 2012.
- [49] A. Sa-ngasoongsong, S. T.S. Bukkapatnam, J. Kim, P. S. Iyer, and R. P. Suresh, "Multi-step sales forecasting in automotive industry based on structural relationship identification," *International Journal of Production Economics*, vol. 140, no. 2, pp. 875–887, 2012, <http://dx.doi.org/10.1016/j.ijpe.2012.07.009>.
- [50] J. Gao, Y. Xie, X. Cui, H. Yu, and F. Gu, "Chinese automobile sales forecasting using economic indicators and typical domestic brand automobile sales data: A method based on econometric model," *Advances in Mechanical Engineering*, vol. 10, no. 2, pp. 1–11, 2018, <http://dx.doi.org/10.1177/1687814017749325>.



# Exploring Levels of ICT Adoption and Sustainable Development – The Case of Polish Enterprises

Ewa Ziemba

University of Economics in Katowice  
1 Maja 50, 40-287 Katowice, Poland  
ewa.ziemba@ue.katowice.pl

**Abstract**— This study is a part research on the effect of information and communication technologies (ICT) adoption on sustainable development in the enterprises' context [1]–[3]. Its main purpose is to identify parameters stimulating the progress of ICT adoption and sustainable development and assess the two constructs based on these parameters. The identified parameters of ICT adoption are grouped into four categories i.e., ICT outlay, information culture, ICT management, and ICT quality, whereas the parameters of sustainable development are classified into ecological, economic, socio-cultural, and political sustainability categories. This study employs a quantitative approach and descriptive statistics are employed to evaluate the levels of ICT adoption and sustainable development. The survey questionnaire was used and data collected from 394 enterprises were analyzed. The research findings reveal that digital and socio-cultural competences of employees and managers, financial capabilities ensuring ICT projects as well as law regulations associated with ICT adoption, and information security were at the highest level within enterprises. However, the lowest level was specific for BI and ERP system adoption as well as the adoption of latest management concepts and the exploitation of synergies between national ICT projects and own ones. Moreover, the improvement of efficiency and effectiveness of customer services, better and more efficient organization of work, the enhancement of customer satisfaction and loyalty and the acquirement of new customers and markets were at the highest level within enterprises. However, the lowest level was specific for enterprises' participation in the democratic public decision-making as well as energy savings and environmental protection. This study advances ongoing research on ICT adoption and sustainable development by exploring parameters which can be used to describe and assess the levels of ICT adoption and sustainable development in the context of enterprises. Moreover, these parameters help clarify areas that need further improvement and stimulate the progress of ICT adoption and sustainable development.

## I. INTRODUCTION

A new paradigm for economic growth, social equality and environmental protection was set in 1987 and introduced the concept of sustainable development to the international community [4]. Sustainable development is a development in which the needs of present generations are met without compromising the chances of future generations to meet their own needs [5]. According to Schauer [6], sustainable development has four dimensions which are ecological, social, economic and cultural sustainability. Ziemba [1] added a political dimension of sustainable development. Furthermore, it can occur at different levels

and within different contexts as many stakeholders on global, national, and community levels are involved in sustainable development [7]. Besides citizens and public administration, enterprises are one of these stakeholders that can contribute to sustainable development and benefit from it [8].

Bisk and Božić [9] highlighted that sustainable development today can best be attained by technological growth, whereas Grunwald [10] assessed the relation between technology and sustainable development as ambivalent. In particular, information and communication technologies (ICT) are a key enabler for sustainable development [11]–[13]. They make significant contributions to revolutionary changes in everyday life, business, and public administration, transforming society and fuelling economic growth. If society stakeholders are unable to acquire the capabilities to adopt ICT effectively, they will be increasingly disadvantaged or even excluded from the benefits afforded by ICT [8]. Some researchers have recognized ICT as one of the most important tools in developing sustainable business practices [14] and supporting the success of businesses [15]. It is contended that ICT enable businesses to improve productivity, foster innovation, cut down costs, increase the effectiveness of processes services, augment the efficiency of business decision-making, react to customer needs at a faster rate, and acquire new ones [16], [14]. Moreover, the ICT adoption by enterprises can gain benefits in environmental preservation by increasing energy efficiency and equipment utilization as well as it can increase information availability to all society stakeholders [6] and as a consequence influence social development [11].

After extensively searching the literature it can be noticed that ICT adoption and sustainable development require in-depth research, inter alia, research on assessing the levels of ICT adoption and sustainable development, and indicating areas that should to be improved. We need to have quantitative tools for describing and measuring the state of ICT adoption and sustainable development in the enterprises' context. These tools should allow to define the direction of desirable actions aimed at facilitating sustainable development as a result of ICT adoption.

There are some indicators and synthetic indexes for assessing ICT adoption, e.g. ICT Development Index (IDI) worked out by International Telecommunication Union [17] and Networked Readiness Index (NRI) of the authorship of

the World Economic Forum [18]. Sets of indicators for measuring sustainable development exist already prepared by Eurostat [19] and OECD [20]. Following an extensive review of the literature, it can be stated that it did not uncover any deep studies providing objective assessment parameters of ICT adoption within enterprises and sustainable development in the enterprises' context. Such parameters would help clarify areas that need further improvement and stimulate the progress of ICT adoption and sustainable development. This paper, therefore, focuses on exploring such parameters. Its aim is to propose parameters describing the progress of ICT adoption and sustainable development and assess the level of ICT adoption and sustainable development based on them.

This paper contributes to the literature in several ways. Section II reviews the current research on ICT adoption within enterprises, sustainable development in the enterprises' context and the assessment of these two constructs. Section III describes the unique research methodology and the data set used for the empirical work. Based on these data, Section IV presents the results, including an assessment of the levels of ICT adoption within enterprises and sustainable development in the enterprises' context. Section V provides the study's contributions, implications, and limitations as well as considerations for future investigative work.

## II. THEORETICAL BACKGROUND AND RESEARCH QUESTION

### A. Sustainable development

There are multiple definitions of the concept of "sustainable development". The most frequently quoted definition comes from the World Commission on Environment and Development, now known as the Brundtland Commission. According to it, the purpose of sustainable development is to meet "the needs of the present without compromising the ability of future generations to meet their own needs" [21, p. 43]. This definition clarifies the primary essence of the concept: the ability to self-sustain development that does not degrade the factors and mechanisms which constitute it.

Looking more closely into the term "sustainable", it is defined as something that is "able to be upheld or defended" [22]. The definition consists of two distinctive parts; the first one implies that sustainable development is the development that can be retained over time, while the second one indicates that sustainable development is the development that can be shielded from the consequences of negative events and processes. These two parts are closely linked, as it is not only events and processes that may affect development, but the means for development may also augment or create new events and fundamental processes that in turn make the task of maintaining development over time exacting [23]. For instance, our dependency of ICT has allowed for great developmental leaps of many societies since the industrial revolution, but is at the same time the

main cause of rising energy consumption and climate change that are now threatening the sole existence of all societies.

The core of the concept of sustainable development embraces two mutually exclusive ideas: the human need to preserve natural resources, and the need to improve the quality of life. Initially, it combined concerns about poverty and development with environmental issues. Then, interpretations of this definition were advanced, ranging from the "pure ecologist" position, through "moderate ecologist," "crash barrier," and "3D," to "4D" [6]. The first two interpretations are purely focused on the ecological dimension. In the "crash barrier" interpretation, the relationship with ecology is weaker, and it places equal weight on social and ecological issues. "3D" defines a further dimension of sustainability, encompassing ecological, social, and economic questions that have equal importance and have to respect each other. In the "4D" approach, cultural dimensions are introduced. Sustainable development is, however, a debatable concept due to its indefinite meaning, which is open to a variety of interpretations, depending upon the given situation [24].

A deeper understanding of the interconnected challenges the world faces allow to recognize that sustainable development has to embrace several sustainability pillars: from the three fundamental pillars related to environmental, economic and social aspects [6], [7] to pillars concerning cultural [25], [26] and political sustainability [25], [27].

This paper defines sustainable development in the enterprises' context as: a dynamic process which enables enterprises to realize their potential and improve their competences and business in ways that simultaneously protect and enhance ecological (Ecl), economic (Eco), socio-cultural (Soc), and political (Pol) sustainability.

Ecological sustainability is the ability of enterprises to retain rates of renewable resource acquisition, pollution creation, and non-renewable resource depletion by means of conservation and appropriate use of air, water, and land resources [28], [29]. Economic sustainability of enterprises means that enterprises can obtain competitive advantage, boost their market share, and increase shareholder value by adopting sustainable practices and models. Among the core drivers of a business case for sustainability are: cost and cost reduction, sales and profit margin, reputation and brand value, innovative capabilities [16], [30]. Socio-cultural sustainability is founded on the socio-cultural aspects that need to be sustained e.g., trust, common meaning, diversity as well as capacity for learning and capacity for self-organization [5]. It is perceived as dependent on social networks, making community contributions, creating a sense of place and offering community stability and security [27], [31]. Political sustainability must be built on the basic values of democracy and effective appropriation of all rights. It is connected with the engagement of enterprises in creating democratic society [27]. Based on the stream of research, Ziemba [1] indicated parameters that fully describe those pillars of sustainable development (Table I).

### B. ICT adoption for sustainable development

ICT represent significant opportunities for sustainable development [11], [12], [32]. The rapid evolution of ICT not only has radically changed of everyday life [33] but also businesses [34]. It has provided enterprises with new instruments to add value to various kinds of sustainability [16], [35].

ICT can be defined as any type of software and hardware used to create, capture, manipulate, communicate, exchange, present, and use information in its various forms” [36, p. 198]. Research on ICT adoption is mainly directed to the development of forecasting studies and the identification of barriers and drivers of technology adoption [37]. Reino et al. [37] indicated two main approaches that can be adopted for the study of ICT adoption phenomenon. These are intra-enterprise and inter-enterprise adoption. The former relates to the process by which ICT are fully adopted by an enterprise from their purchase to the full integration as part of the business strategy. The latter refers to the phenomena by which ICT adoption take place among an enterprise and its stakeholders as consumers, public administration and other enterprises [38]. It should therefore be acknowledged that ICT adoption takes place in stages and this implies that different levels of ICT adoption can be identified within enterprises.

Furthermore, many investigators, as well as developmental organizations recognize the significance of ICT for sustainable development [6], [11], [35], [39], [40]. ICT are accelerators, amplifiers, and augmenters of sustainable development. They make it feasible to enhance sustainable development more flexibly and dynamically. More pointedly, ICT presents opportunities to make trade-offs

between economic growth, the environment and social cohesion as well as culture and political issues [41]. Equally, there is the opportunity to maximize the social, ecological, economic and cultural opportunities of ICT and mitigate its adverse impacts.

In this study, ICT adoption has been explored in terms of intra-enterprise. Nevertheless, some issues of inter-enterprise adoption have been taken into consideration e.g., related to an enterprise’s collaboration with its customers. ICT adoption is understood as the whole spectrum of activities from the period when enterprises justify the need for adopting ICT until the period when enterprises experience the full potential of ICT and derive ecological, economic, socio-cultural and political sustainability from them [1]. The following four pillars of ICT adoption within enterprises are recognized: ICT outlay (Out), information culture (Cul), ICT management (Man), and ICT quality (Qua) [1], [2].

ICT outlay consists of the enterprises’ financial capabilities and expenditure on the ICT adoption, as well as funding acquired by enterprises from the European funds. The information culture component encompasses digital and socio-cultural competences of enterprises’ employees and managers, constant enhancement of these competences, personal mastery, and incentive systems fostering ICT adoption by employees. The ICT management component embraces the alignment between business and ICT, top management support for ICT projects in the entire ICT adoption lifecycle, implementation of law regulations associated with the ICT adoption, regulations on ICT and information security and protection. The ICT quality component comprises the quality and security of back- and front-office information systems, quality of hardware,

TABLE I.  
PARAMETERS OF ICT ADOPTION AND SUSTAINABLE DEVELOPMENT IN THE ENTERPRISES’ CONTEXT

Parameters of ICT adoption				Parameters of sustainable development	
<b>Out1</b>	Financial capabilities	<b>Man16</b>	ICT project team	<b>Ecl1</b>	Sustainability in ICT
<b>Out2</b>	Expenditure on ICT	<b>Man17</b>	Top management support	<b>Ecl2</b>	Sustainability by ICT
<b>Out3</b>	Funding acquired from the European funds	<b>Man18</b>	Management concepts adoption	<b>Eco3</b>	Cost reduction
<b>Cul4</b>	Managers’ ICT competences	<b>Man19</b>	Information security regulations	<b>Eco4</b>	Sales growth
<b>Cul5</b>	Employees’ ICT competences	<b>Man20</b>	ICT regulations	<b>Eco5</b>	Product development
<b>Cul6</b>	Managers’ permanent education	<b>Man21</b>	ICT public project	<b>Eco6</b>	Effective and efficient management
<b>Cul7</b>	Employees’ permanent education	<b>Man22</b>	Competitive ICT market	<b>Eco7</b>	Effective and efficient customer service
<b>Cul8</b>	Employees’ personal mastery	<b>Qua23</b>	ICT infrastructure quality	<b>Eco8</b>	Effective and efficient work
<b>Cul9</b>	Managers’ socio-cultural competences	<b>Qua24</b>	Back-office system quality	<b>Eco9</b>	Acquiring new customers and markets
<b>Cul10</b>	Employees’ socio-cultural competences	<b>Qua25</b>	Front-office system quality	<b>Eco10</b>	Increasing customer satisfaction/loyalty
<b>Cul11</b>	Employees’ creativity	<b>Qua26</b>	Back-office system security	<b>Soc11</b>	Competence extension
<b>Cul12</b>	Incentive systems	<b>Qua27</b>	Front-office system security	<b>Soc12</b>	Working environment improvement
<b>Man13</b>	Alignment between business strategy and ICT	<b>Qua28</b>	E-service maturity levels	<b>Soc13</b>	Increasing security
<b>Man14</b>	Supporting business models by ICT	<b>Qua29</b>	ERP adoption	<b>Soc14</b>	Reducing social exclusion
<b>Man15</b>	ICT management procedure	<b>Qua30</b>	BI (Business Intelligence) adoption	<b>Pol15</b>	E-democracy
---	---	---	---	<b>Pol16</b>	E-public services

Source: on the basis of [22].

maturity of e-services, and adoption of ERP and BI systems. Table I describes each of the above ICT adoption pillars.

### C. Problem identification and research questions

ICT adoption for sustainable development is not a destination, but a dynamic process of adaptation, learning and action. It is about recognizing, understanding and acting on ICT adoption and sustainable development as well as on interconnections between them.

As mentioned above, in the previous study Ziemia [1] indicated parameters describing the constructs of ICT adoption and sustainable development in the enterprises' context, and then grouped them into appropriate pillars (Table I). ICT adoption embraces ICT outlay, information culture, ICT management and ICT quality, whereas sustainable development includes ecological, economic, socio-cultural and political sustainabilities. Then the quality of the two constructs was assessed by examining the construct reliability [42], convergent validity [43], [44], and discriminant validity [43], [45]. Overall, the results successfully established the reliability as well as convergent and discriminant validity of ICT adoption and sustainable development, and their pillars. Furthermore, the levels of ICT and sustainable development pillars were assessed (Table II) and the approach to the measurement of the two constructs ICT was proposed [3].

## III. RESEARCH METHODOLOGY

To address the main research problem and answer the research questions a quantitative research approach was adopted. Research methods included a critical review of the literature, logical deduction, a survey questionnaire, and statistical analysis. The research process has been described in the previous works [1], [2] but for the ease of this paper readability and sake of its completeness, it is also presented below.

### A. Research instrument

The Likert-type instrument (survey questionnaire) was developed. Closed-ended questions were specified to collect data regarding the evaluation of the parameters describing:

- The four pillars of ICT adoption i.e., ICT outlay (Out), information culture (Cul), ICT management (Man), and ICT quality (Qua) (Table I). The respondents answered the question: *Using a scale of 1 to 5, state to what extent do you agree that the following situations and phenomena result in the efficient and effective ICT adoption in your enterprise?* The scale's descriptions were: 5 – strongly agree, 4 – rather agree, 3 – neither agree nor disagree, 2 – rather disagree, 1 – strongly disagree; and
- The four pillars of sustainable development i.e., ecological (Ecl), economic (Eco), socio-cultural (Soc),

TABLE II.  
THE LEVELS OF ICT ADOPTION AND SUSTAINABLE DEVELOPMENT IN THE ENTERPRISES' CONTEXT (N=394)

Pillar	Mean	Q25	MDN	Q75	VAR	SD	CV in %	SK	CK
<b>ICT adoption pillars</b>									
<b>Out</b>	3.78	3.33	4.00	4.33	0.71	0.84	22.33	-0.78	0.35
<b>Cul</b>	3.71	3.22	3.78	4.33	0.57	0.75	20.32	-0.46	-0.35
<b>Man</b>	3.58	3.10	3.60	4.20	0.62	0.79	22.07	-0.55	-0.17
<b>Qua</b>	3.60	3.00	3.75	4.25	0.74	0.86	23.95	-0.56	-0.22
<b>Sustainable development pillars</b>									
<b>Ecl</b>	3.44	3.00	3.50	4.00	1.03	1.01	29.48	-0.40	-0.58
<b>Eco</b>	3.68	3.25	3.75	4.25	0.62	0.79	21.38	-0.78	0.65
<b>Soc</b>	3.51	3.00	3.75	4.25	0.78	0.88	25.14	-0.46	-0.35
<b>Pol</b>	3.44	3.00	3.50	4.00	1.02	1.01	29.41	-0.47	-0.47

Note: mean, median (MDN), first quartile (Q25), third quartile (Q75), variance (VAR), standard deviation (SD), coefficient of variation (CV), skewness (SK), and coefficient of kurtosis (CK).

Source: [2]

The present study examines and evaluates particular parameters shaping each of ICT adoption and sustainable development pillars in the context of Polish enterprises. It focuses on addressing the following two research question:

RQ1: What is the level of ICT adoption in Polish enterprises?

RQ2: What is the level of sustainable development in the context of Polish enterprises?

and political sustainability (Pol) (Table I). The respondents answered the question: *Using a scale of 1 to 5, evaluate the following benefits for your enterprise resulting from the efficient and effective ICT adoption?* The scale's descriptions were: 5 – strongly large, 4 – rather large, 3 – neither large nor disagree, 2 – rather small, 1 – strongly small.

### B. Research subjects and procedure

In April 2016, the pilot study was conducted to verify the survey questionnaire. Ten experts participated in the study i.e., five researchers in business informatics and five managers from five enterprises – leaders in the ICT application. Finishing touches were put into the questionnaire, especially of a formal and technical nature. No substantive amendments were required.

The subjects in the study were enterprises from the Silesian Province in Poland. The choice of this region was driven by the fact of its continuous and creative transformations related to restructuring and reducing the role of heavy industry in the development of research and science, supporting innovation, using *know-how* and transferring new technologies, as well as increasing importance of services. In response to the changing socio-economic and technological environment intensive work on the development of the information society has been undertaken in the region for several years. In the next development strategies of the information society it was and is assumed that the potential of the region, especially in the design, provision and use of advanced information and communication technologies will be increased [46]. All this means that the results of this research can be reflected in innovative efforts to build a sustainable information society in the region and, at the same time, constitute *a modus operandi* for other regions throughout the country and other countries.

Selecting a sample is a fundamental element of a positivistic study [47]. The stratified sampling and snowball sampling were therefore used to obtain the sample that can be taken to be true for the whole population. The strata were identified based on enterprise's size (defined in terms of the number of employees), economy sector, and type of business activity (defined in terms of related to ICT and non-ICT activities).

The subjects were advised that their participation in completing the survey was voluntary. At the same time, they were assured anonymity and guaranteed that their responses would be kept confidential.

### C. Data collection

Having applied the Computer Assisted Web Interview and employed the SurveyMonkey platform, the survey questionnaire was uploaded to the website. The data were collected during a two-month period of intense work, between May 12, 2016 and July 12, 2016. After screening the responses and excluding outliers, there was a final sample of 394 usable, correct, and complete responses. The sample error for an infinite population was of about 5% for a confidence level 97% ( $p = q = 0.5$ ) which previous studies have suggested as acceptable [48], [49]. Additionally, it presented a successful representation of the different business types, economy sectors and size categories.

Table III provides details about enterprise's size, type of the business activities, and economy sector.

TABLE III.  
ANALYSIS OF ENTERPRISES PROFILES (N=394)

Characteristics	Frequency	Percentage
<b>Number of employees</b>		
250 and above (large)	78	19.80%
50–249 (medium)	83	21.07%
10–49 (small)	122	30.96%
less than 10 (micro)	111	28.17%
<b>Economy sector</b>		
I sector – producing raw material and basic foods	27	6.85%
II sector – manufacturing, processing, and construction	83	21.07%
III sector – providing services to the general population and to businesses	238	60.40%
IV sector – including intellectual activities	46	11.68%
<b>Business activities</b>		
ICT (manufacturing, trade, services)	136	34.52%
No ICT	258	65.48%

Source: own elaboration.

### D. Data analysis

The data were stored in Microsoft Excel format. Using Statistica package and Microsoft Excel, the data were analyzed. The descriptive statistical analysis was employed to describe the levels of ICT adoption and sustainable development parameters within enterprises. The following statistics were calculated: mean, median (MDN), first quartile (Q25), third quartile (Q75), mode, variance (VAR), standard deviation (SD), coefficient of variation (CV), skewness (SK), and coefficient of kurtosis (CK).

## IV. RESEARCH FINDINGS

### A. The level of ICT adoption within enterprises

In order to answer the research question *RQ1: What is the level of ICT adoption in Polish enterprises?*, a detailed descriptive analysis was conducted. The results are presented in Table IV.

It has been found that the average levels of ICT adoption parameters ranged from 3.24 to 4.22 (on a 5-point scale from 1.00 to 5.00). The median values were in the range between 3.00 and 5.00, whereas the mode values were 4 or 5. On average, the highest levels are specific for parameters related mainly to three ICT adoption pillars i.e., information culture, ICT management, and ICT outlay.

The highest ranked parameters of ICT adoption were (Table IV):

TABLE IV.  
THE LEVELS OF ICT ADOPTION PARAMETERS IN THE ENTERPRISES' CONTEXT (N=394)

Parameters	Mean	Q25	MDN	Q75	Mode	Sample volume for Mode	VAR	SD	CV in %
Out1	3.98	4	4	5	4	158	1.18	1.09	27.27
Out2	3.71	3	4	5	4	165	1.27	1.13	30.37
Out3	3.64	3	4	5	4	132	1.54	1.24	34.14
Cul4	4.20	4	5	5	5	207	1.16	1.08	25.62
Cul5	4.22	4	4	5	5	186	0.92	0.96	22.73
Cul6	3.53	2	4	5	4	127	1.62	1.27	36.12
Cul7	3.47	2	4	4	4	142	1.58	1.26	36.26
Cul8	3.28	2	3	4	4	139	1.30	1.14	34.77
Cul9	3.85	3	4	5	4	170	1.22	1.10	28.73
Cul10	3.91	4	4	5	4	192	0.98	0.99	25.34
Cul11	3.58	3	4	4	4	188	1.21	1.10	30.70
Cul12	3.37	2	4	4	4	142	1.57	1.25	37.16
Man13	3.53	3	4	4	4	165	1.27	1.13	32.00
Man14	3.61	3	4	4	4	181	1.11	1.05	29.18
Man15	3.60	3	4	4	4	152	1.43	1.20	33.28
Man16	3.52	2	4	5	4	145	1.64	1.28	36.40
Man17	3.68	3	4	4	4	171	1.25	1.12	30.36
Man18	3.47	3	4	4	4	157	1.36	1.17	33.60
Man19	3.88	3	4	5	4	160	1.27	1.13	29.05
Man20	3.92	3	4	5	4	148	1.24	1.11	28.36
Man21	3.25	2	3	4	4	122	1.60	1.27	38.93
Man22	3.33	2	3	4	4	125	1.40	1.18	35.48
Qua23	3.66	3	4	5	4	151	1.49	1.22	33.34
Qua24	3.68	3	4	4	4	170	1.25	1.12	30.43
Qua25	3.71	3	4	5	4	161	1.25	1.12	30.10
Qua26	3.77	3	4	5	5	132	1.44	1.20	31.78
Qua27	3.75	3	4	5	4	142	1.50	1.22	32.67
Qua28	3.53	3	4	4	4	160	1.42	1.19	33.73
Qua29	3.44	2	4	4	4	130	1.57	1.25	36.52
Qua30	3.24	2	4	4	4	134	1.69	1.30	40.20

Note: mean, median (MDN), first quartile (Q25), third quartile (Q75), variance (VAR), standard deviation (SD), coefficient of variation (CV), skewness (SK), and coefficient of kurtosis (CK).

- Cul5 (mean = 4.22, MDN = 4, mode = 5) and Cul4 (mean = 4.20, MDN = 5, mode = 5). It means that digital competences of enterprises' employees and managers are relatively high. Managers and employees are able to operate a computer and the Internet, use a different kind of software and applications, search for information, use it and evaluate its usefulness, as well as creatively, efficiently and effectively use ICT so as to achieve a variety of business benefits;
- Cul10 (mean = 3.91, MDN = 4, mode = 5) and Cul9 (mean = 3.85, MDN = 4, mode = 4). It means that socio-cultural competences of enterprises' employees and managers are also relatively high. Managers and employees are open to change and novelties, can negotiate, integrate the team and build confidence, are able to manage a group as well as a multicultural team, know how to build varied relationships and networks, share knowledge and are able to manage knowledge;
- Out1 (mean = 3.98, MDN = 4, mode = 4). It means that enterprises' financial capabilities ensure the purchase and use of computer hardware, software, the Internet, telecommunications and improvement of digital literacy;
- Man20 (mean = 3.92, MDN = 4; mode = 4). It means that enterprises implement and apply the law regulations associated with ICT adoption, in particular related to electronic invoicing, electronic signatures, data protection, electronic services, protection of databases, distance contracts;
- Man19 (mean = 3.88, MDN = 4; mode = 4). It means that enterprises develop and apply regulations and tools

on information security and protection of personal data, also associated with ICT adoption in terms of intra-enterprise and inter-enterprise (e.g., in relations with customers and business partners).

Furthermore, the lowest levels of ICT adoption were mainly related to ICT quality and ICT management (Table IV):

- Qua30 (mean = 3.24, MDN = 4, mode = 4). It means that many enterprises did not implement any BI system and do not employ any business analyses e.g., sales, customers, financial, marketing, products analyses. In general, 25% of enterprises assessed BI adoption at a level not higher than 2.00 and 75% of enterprises – at a level not higher than 4.00. Coefficient of variation with the value above 40% shows substantial differences in BI adoption within enterprises.
- Qua29 (mean = 3.44, MDN = 4, mode = 4). It means that many enterprises did not implement any ERP system(or any integrated domain-specific systems) which provides a coherent, comprehensive and integrated support for business processes in the whole range of business activities, and supports the primary and secondary business processes, such as sales, purchasing, marketing, distribution, customer service, warehouse management, human resources and payroll, or finance and accounting;
- Man21 (mean = 3.25, MDN = 3, mode = 4) concerning the implementation and apply of ICT within enterprises arising from ICT public projects, coordinated and

implemented at national, regional and/or local level e.g., the construction of broadband networks, making electronic platforms of public services available etc.;

- Man18 (mean = 3.47, MDN = 4, mode = 4). It means that a lot of enterprises did not implement the latest management concepts, such as process approach, knowledge management, risk management, change management, quality management, customer relationship management, trust management, human resource management, networking approach.

In general, the level of ICT outlay was the highest within enterprises, followed by the level of information culture. The levels of ICT management and ICT quality were the lowest (Table II).

#### B. The level of sustainable development in the enterprises' context

In order to answer the research question *RQ2: What is the level of sustainable development in the context of Polish enterprises?*, a detailed descriptive analysis was conducted. The results are presented in Table V.

It has been found that the average levels of sustainable development parameters ranged from 3.25 to 3.96 (on a 5-point scale from 1.00 to 5.00). The median and mode values were 4.00 except for Pol15 with the MDN = 3.00. On average, the highest levels are mainly specific for parameters related to economic sustainability, whereas a parameter of political sustainability was ranked the lowest. The highest ranked parameters of sustainability were (Table V):

TABLE V.  
THE LEVELS OF SUSTAINABLE DEVELOPMENT PARAMETERS IN THE ENTERPRISES' CONTEXT (N=394)

Parameters	Mean	Q25	MDN	Q75	Mode	Sample volume for Mode	VAR	SD	CV in %
Ecl1	3.38	2	4	4	4	149	1.21	1.10	32.53
Ecl2	3.50	2	4	4	4	160	1.40	1.18	33.82
Eco3	3.48	3	4	4	4	166	1.24	1.11	31.99
Eco4	3.67	3	4	4	4	168	1.11	1.05	28.70
Eco5	3.54	3	4	4	4	163	1.36	1.17	32.95
Eco6	3.51	3	4	4	4	159	1.31	1.14	32.62
Eco7	3.96	4	4	5	4	180	1.01	1.00	25.35
Eco8	3.89	4	4	5	4	185	1.03	1.02	26.09
Eco9	3.68	3	4	5	4	150	1.27	1.13	30.64
Eco10	3.74	3	4	5	4	165	1.14	1.07	28.50
Soc11	3.76	3	4	4	4	181	1.01	1.00	26.74
Soc12	3.45	2	4	4	4	144	1.52	1.23	35.70
Soc13	3.47	2	4	4	4	150	1.50	1.23	35.29
Soc14	3.38	2	4	4	4	153	1.26	1.12	33.29
Pol15	3.25	2	3	4	4	148	1.32	1.15	35.40
Pol16	3.63	3	4	4	4	178	1.24	1.11	30.61

Note: mean, median (MDN), first quartile (Q25), third quartile (Q75), variance (VAR), standard deviation (SD), coefficient of variation (CV), skewness (SK), and coefficient of kurtosis (CK).

- Eco7 (mean = 3.96, MDN = 4, mode = 4). It means that the improvement of efficiency and effectiveness of customer services resulting of ICT adoption was evaluated relatively high by enterprises;
- Eco8 (mean = 3.96, MDN = 4, mode = 4). It means that thanks to ICT adoption the enterprise achieves better and more efficient organization of work resulting from improvements and automation of business processes, communication, collaboration and networking within the enterprise and in its relations with its stakeholders (customers, suppliers, partners), facilitating access to information;
- Eco10 (mean = 3.74, MDN = 4, mode = 4). It means that the improvement of customer satisfaction and loyalty from products and services offered to them by the enterprise as well as pre- and post-sales support resulting from ICT adoption was assessed relatively high by enterprises;
- Eco9 (mean = 3.68, MDN = 4, mode = 4). It means that the result of ICT adoption by enterprise is to acquire new customers and markets, including foreign ones e.g., through internet marketing, online sales, obtaining information on markets and customers;
- Soc11 (mean = 3.76, MDN = 4, mode = 4). It means that ICT adoption by enterprise allows to extend knowledge and skills already held by employees and acquire new ones (including digital knowledge and skills), as well as better align thinking and action in response to the changing reality, legal requirements and customer needs.

Furthermore, the lowest level of sustainability was related to political, ecological, and socio-cultural sustainability (table V):

- Pol15 (mean = 3.25, MDN = 4, mode = 4). It means that enterprises' participation in the public consultation and democratic public decision-making as well as development of cooperation, communication, partnerships and networks between enterprises and public administration were assessed relatively very low;
- Ec11 (mean = 3.38, MDN = 4, mode = 4). It means that a lot of enterprises did not achieve lower average annual energy consumption and increased protection of the environment through ICT consuming less energy and built with fewer materials (miniaturization), and more easily recyclable and disposable;
- Soc14 (mean = 3.38, MDN = 3, mode = 4). It means that reducing social exclusion due to age, education, place of residence or disability, by facilitating access to the enterprise, its products/services and jobs was ranked relatively very low.

On average, the level of economic sustainability was the highest, whereas the levels of ecological and political sustainability were the lowest (Table II).

## V. CONCLUSIONS

### A. Research contribution

Although the literature review suggested that the phenomenon of ICT adoption for sustainable development had been previously examined [6], [11], [12], [16], [32], [35], this study extended previous research on the contribution of ICT adoption by enterprises to sustainable development [1]–[3]. and indicated the levels of ICT adoption and sustainable development in-depth. It contributes to the existing research on sustainable information society, ICT adoption, and sustainable development, in particular in the enterprises' context by:

- indicating and assessing the level of ICT adoption, especially in terms of ICT outlay, information culture, ICT management, and ICT quality; and
- indicating and assessing the level of sustainable development, especially in terms of ecological, economic, socio-cultural, and political sustainability.

Firstly, this study indicated that ICT outlay was at the highest level followed by information culture, whereas the lowest and similar levels were specific to ICT management and ICT quality. Digital and socio-cultural competences of employees and managers, financial capabilities ensuring any ICT projects as well as law regulations associated with ICT adoption and information security were relatively highly ranked by enterprises. However, the lowest level was specific for BI and ERP system adoption as well as the adoption of latest management concepts and exploitation of synergies between national ICT projects and own ones. All these require to improve ICT adoption, mainly its quality and management pillars.

Secondly, the outcomes showed that economic sustainability was at the highest level, whereas the lowest and similar levels were specific to ecological and political sustainability. The improvement of efficiency and effectiveness of customer services, better and more efficient organization of work, the enhancement of customer satisfaction and loyalty as well as the increase of new customers and markets as a result of ICT adoption were evaluated relatively high by enterprises. However, the lowest level was specific for enterprises' participation in the democratic public decision-making as well as energy savings and environment protection were assess relatively very low. It means that enterprises reap more economic benefits than ecological and political ones from adopting ICT. It is, therefore, required to increase ecological and political sustainability through ICT adoption.

### B. Research implication for research and practice

While this research is exploratory, it should provide a valuable foundation for further work examining ICT adoption, sustainable development, and a synergy between them more widely.

Researchers may use the proposed methodology to do similar analyses with different sample groups in other countries, and many comparisons between different countries can be drawn. Moreover, the methodology constitutes a very comprehensive basis for identifying the levels of ICT adoption and sustainable development, as well as the correlations between the two constructs, but researchers may develop, verify and improve this methodology.

This study offers several implications for enterprises. They may find the results appealing and useful in enhancing ICT adoption, experiencing the full potential of ICT adoption, and deriving various benefits from ICT adoption. The results suggest various kinds of advantages like ecological, economic, socio-cultural, and political that can be gained thanks to ICT adoption. In addition, they propose some guidelines on how to effectively and efficiently adopt ICT in order to obtain those advantages. It is evident from the findings that Polish enterprises should devote utmost attention to the enhancement of ICT management and ICT quality. Most of all, this research can be genuinely useful for the transition economies in Central and Eastern Europe. This is because the countries are similar with regard to analogous geopolitical situation, their joint history, traditions, culture and values, the quality of ICT infrastructure, as well as developing democratic state structures and a free-market economy, and participating in the European integration process.

All in all, the research results might provide a partial explanation to the issue of how enterprises can participate in the creation of sustainable development.

### C. Research limitations and future works

However, as with many other studies, this research looking more than superficially into ICT adoption and sustainable development in the enterprises' context has been limited. First, the ICT adoption and sustainability constructs are new constructs that have yet to be further explored and exposed to repeated empirical validation. Second, the sample consisted of Polish enterprises only, especially from the Silesian Province. The study sample excludes statistical generalization of the results from Silesian enterprises to Polish enterprises. However, previous research into the success factors for and the level of adopting ICT in Poland [50] indicated that there is no difference between enterprises in the Silesia Province and in Poland. Therefore, these research findings cannot be confined only to the Silesian enterprises and can be extended to Polish enterprises. After all, caution should be taken when generalizing the findings to other regions and countries. Finally, the research subjects were limited to enterprises and it is therefore only the standpoint of enterprises toward ICT adoption for achieving sustainable development. Caution should be taken when generalizing the findings to sustainable development in general.

Additional research must be performed to better understand ICT adoption and sustainable development. First, further validation of the levels of ICT adoption and sustainable development should be carried out for a larger sample comprising enterprises from different Polish provinces as well as from other countries. Second, research on the measurement of ICT adoption and sustainability in households and government units should be conducted because they are, besides enterprises, the main stakeholders of SIS.

### REFERENCES

- [1] E. Ziemia, "The contribution of ICT adoption to the sustainable information society," *Journal of Computer Information Systems*, vol. 59, issue 2, pp. 116–126, 2019, <https://doi.org/10.1080/08874417.2017.1312635>.
- [2] E. Ziemia, "The ICT adoption in enterprises in the context of the sustainable information society," in *Proceedings of the 2017 Federated Conference on Computer Science and Information Systems FedCSIS*, M. Ganzha, L. Maciaszek, and M. Paprzycki, Eds., Czech Technical University in Prague, Prague, September 3-6, 2017, p. 1031–1038. <https://doi.org/10.15439/2017F89>.
- [3] E. Ziemia, "Synthetic indexes for a sustainable information society: Measuring ICT adoption and sustainability in Polish enterprises," in *Information technology for management: Ongoing Research and Development*, E. Ziemia, Ed. *Lecture Notes in Business Information Processing*, vol. 311, pp. 151–169, 2018. [https://doi.org/10.1007/978-3-319-77721-4\\_9](https://doi.org/10.1007/978-3-319-77721-4_9).
- [4] S. Sala, F. Farioli, and A. Zamagni, "Progress in sustainability science: lessons learnt from current methodologies for sustainability assessment: Part 1," *The International Journal of Life Cycle Assessment*, vol. 18, no 781, pp. 1653–1672, 2012. <https://doi.org/10.1007/s11367-012-0508-6>.
- [5] M. Missimer, K.H. Robèrt, and G. Broman, "A strategic approach to social sustainability-Part 2: A principle-based definitions," *Journal of Cleaner Production*, vol. 149, no 1, pp. 42–52, 2017. <https://doi.org/10.1016/j.jclepro.2016.04.059>.
- [6] T. Schauer, *The sustainable information society – vision and risks*. Vienna: The Club of Rome – European Support Centre, 2003.
- [7] J. Servaes and N. Carpentier, Eds. *Towards a sustainable information society. Deconstructing WSIS*. Portland: Intellect, 2006.
- [8] E. Ziemia, Eds. *Towards a sustainable information society: People, business and public administration perspectives*. Newcastle upon Tyne: Cambridge Scholars Publishing, 2016.
- [9] T. Bisk and P. Božtuć, "Sustainability as growth," in *Technology, society and sustainability. Selected concepts, issues and cases*, ed. L.W. Zacher, Ed. Cham: Springer, 2017, pp. 239–250. [https://doi.org/10.1007/978-3-319-47164-8\\_16](https://doi.org/10.1007/978-3-319-47164-8_16).
- [10] A. Grunwald, "Technology assessment and policy advice in the field of sustainable development," in *Technology, society and sustainability. Selected concepts, issues and cases*, L.W. Zacher, Ed. Cham: Springer, 2017, pp. 203–221. [https://doi.org/10.1007/978-3-319-47164-8\\_14](https://doi.org/10.1007/978-3-319-47164-8_14).
- [11] L.M. Hilty and B. Aebischer, "ICT for sustainability: An emerging research field," *Advances in Intelligent Systems and Computing*, vol. 310, pp. 1–34, 2015.
- [12] L.M. Hilty and M.D. Hercheui, "ICT and sustainable development, What kind of information society?," in *What kind of information society? Governance, virtuality, surveillance, sustainability, resilience, Proceedings of 9th IFIP TC 9 International Conference, HCC9, and 1st IFIP TC 11 International Conference*, J. Berleur, M.D. Hercheui, and L.M. Hilty, Eds., Brisbane. September 20-23, 2010, p. 227–235.
- [13] J.W. Houghton, "ICT and the environment in developing countries: A review of opportunities and developments," in *What kind of information society? Governance, virtuality, surveillance, sustainability, resilience, Proceedings of 9th IFIP TC 9 International Conference, HCC9, and 1st IFIP TC 11 International Conference*,

- J. Berleur, M.D. Hercheui, and L.M. Hilty, Eds., Brisbane, September 20-23, 2010, p. 236–247.
- [14] R.T. Watson, M.C. Boudreau, A.J. Chen, and M. Huber, “Green IS: Building sustainable business practices,” in *Information systems*, R.T. Watson, Ed. Athens: Global Text Project, 2008, pp. 247–261.
- [15] V. Kodakanchi, E. Abueyaman, M.H.S. Kuofie, and J. Qaddour, “An economic development model for IT in developing countries,” *The Electronic Journal of Information Systems in Developing Countries*, vol. 28, no 7, pp. 1–9, 2006.
- [16] M.G. Guillemette and G. Paré, “Toward a new theory of the contribution of the IT function in organizations,” *MIS Q*, (36:2), 2012, pp. 529–551.
- [17] ITU, *Measuring the Information Society Report*. Geneva: International Telecommunication Union, 2017. [https://www.itu.int/en/ITU-D/Statistics/Documents/publications/misr2017/MISR2017\\_Volume1.pdf](https://www.itu.int/en/ITU-D/Statistics/Documents/publications/misr2017/MISR2017_Volume1.pdf) (accessed: 12th April 2019).
- [18] WEF, *Networked Readiness Index*. World Economic Forum, 2019. <http://reports.weforum.org/global-information-technology-report-2016/networked-readiness-index/> (accessed: 12th April 2019).
- [19] Eurostat, *Sustainable development in the European Union. Monitoring report on progress. Towards the SDGS in an EU context*. Luxembourg: Publications Office of the European Union, 2018. <https://ec.europa.eu/eurostat/documents/3217494/9237449/KS-01-18-656-EN-N.pdf/2b2a096b-3bd6-4939-8ef3-11cfc14b9329> (accessed: 12th April 2019).
- [20] OECD, *Measuring distance to the SDG targets. An assessment of where OECD countries stand*. OECD, 2017. <https://www.oecd.org/sdd/OECD-Measuring-Distance-to-SDG-Targets.pdf> (accessed: 12th April 2019).
- [21] WCED, *Our common future*. New York: Oxford University Press, 1987.
- [22] *New Oxford American Dictionary*, <https://en.oxforddictionaries.com/definition/sustainable> (accessed: 12th April 2019).
- [23] P. Becker, *Sustainability Science. Managing Risk and Resilience for Sustainable Development*. Elsevier B.V., 2014. <https://doi.org/10.1016/B978-0-444-62709-4.00005-1>.
- [24] M. Zemigala, “Tendencies in research on sustainable development in management sciences,” *Journal of Cleaner Production*, vol. 218, pp. 796–809, 2019. <https://doi.org/10.1016/j.jclepro.2019.02.009>.
- [25] R. Axelsson, P. Angelstam, E. Degerman, S. Teitelbaum, K. Andersson, M. Elbakidze, and M.K. Drotz, “Social and cultural sustainability: Criteria, indicators, verifier variables for measurement and maps for visualization to support planning,” *AMBIO*, vol. 42, pp. 215–228, 2013. <https://doi.org/10.1007/s13280-012-0376-0>.
- [26] K. Soini and J. Deseine J. “Culture-sustainability relation: Towards a conceptual framework,” *Sustainability*, vol. 8, no. 2, paper 167, 2016. <https://doi.org/10.3390/su8020167>.
- [27] R. Khan, “How frugal innovation promotes social sustainability,” *Sustainability*, vol. 8, no 10, paper 1034, 2016. <https://doi.org/10.3390/su8101034>
- [28] A.H. Huang, “A model for environmentally sustainable information systems development,” *Journal of Computer Information Systems*, vol. 49, no 4, pp. 114–121, 2009.
- [29] B. Moldan, S. Janoušková, and T. Hák, “How to understand and measure environmental sustainability: Indicators and targets,” *Ecological Indicators*, vol. 17, pp. 4–13, 2012.
- [30] M.G. Guillemette and G. Paré, “Transformation of the information technology function in organizations: A Case study in the manufacturing sector,” *Canadian Journal of Administrative Sciences*, vol. 29, pp. 177–190, 2012.
- [31] T. Hameed, *ICT as an enabler of socio-economic development*. Daejeon: Information & Communications University, 2015, <http://www.itu.int/osg/spu/digitalbridges/materials/hameed-paper.pdf>, (accessed: 12th June 2016).
- [32] B. Donnellan, C. Sheridan, and E. Curry, “A capability maturity framework for sustainable information and communication technology,” *IT Professional*, vol. 13, no 1, pp. 33–40, 2011.
- [33] P. Palvia, N. Baqir, and H. Nemati, “ICT for socio-economic development: A citizens’ perspective,” *Information & Management*, vol.55, pp. 160–176, 2018. <https://doi.org/10.1016/j.im.2017.05.003>
- [34] N. Roztocki and H.R. Weistroffer, “Conceptualizing and researching the adoption of ict and the impact on socioeconomic development,” *Information Technology for Development*, vol. 22, no. 4, pp. 541–549, 2016. <http://dx.doi.org/10.1080/02681102.2016.1196097>
- [35] E. Curry and B. Donnellan, “Understanding the maturity of sustainable ICT,” in *Green business process management – Towards the sustainable enterprise*, J. vom Brocke, S. Seidel, and J. Recker, Eds. Berlin: Springer, 2012, pp. 203–216.
- [36] R. Ryssel, T. Ritter, and H.G. Gemunden, “The impact of information technology deployment on trust, commitment and value creation in business relationships,” *Journal of Business & Industrial Marketing*, vol. 19, no. 3, pp. 197–207, 2004.
- [37] S. Reino, A.J. Frew, and C. Albacete-Sáez, “ICT adoption and development: issues in rural accommodation,” *Journal of Hospitality and Tourism Technology*, vol. 2, issue 1, pp. 66–80, 2011. <https://doi.org/10.1108/17579881111112421>
- [38] G. Battisti, and P. Stoneman, “Inter- and intra-firm effects in the diffusion of new process technology,” *Research Policy*, vol. 32, no. 9, pp. 1641–55, 2003.
- [39] T. Niebel, “ICT and economic growth – Comparing developing, emerging and developed countries,” *World Development*, vol. 104, pp. 197–211, 2018. <https://doi.org/10.1016/j.worlddev.2017.11.024>.
- [40] K.M. Vu, 2013. Information and communication technology (ICT) and Singapore’s economic growth,” *Information Economics and Policy*, vol. 25, pp. 284–300, 2013. <http://dx.doi.org/10.1016/j.infoecopol.2013.08.002>
- [41] EITO, *The impact of ICT on sustainable development*. European Information Technology Observatory and Forum for the Future, 2002. [http://homepage.cs.latrobe.edu.au/sloke/greenIT/eito\\_forum\\_2002.pdf](http://homepage.cs.latrobe.edu.au/sloke/greenIT/eito_forum_2002.pdf) (accessed: 12th April 2019).
- [42] P.R. Hinton, C. Brownlow, I. McMurvay, and B. Cozens, *SPSS Explained*. East Sussex: Routledge, 2004.
- [43] D. Gefen and D. Straub, “A practical guide to factorial validity using PLS-graph: Tutorial and annotated example,” *Communications of the Association for Information Systems*, vol. 16, no 5, pp. 91–109, 2005.
- [44] J. Hulland, “Use of Partial Least Squares (PLS) in strategic management research: A review of four recent studies,” *Strategic Management Journal*, vol. 20, no 2, p. 195–204, 1999.
- [45] T.A. Brown, *Confirmatory factor analysis for applied research*. Guilford Press, 2006.
- [46] ŚCSI, *Strategia rozwoju społeczeństwa informacyjnego województwa śląskiego do roku 2015 [Strategy of information society development in Upper Silesia region]*. Katowice: Śląskie Centrum Społeczeństwa Informacyjnego, 2009, [http://www.e-slask.pl/article/strategia\\_rozwoju\\_spoleczenstwa\\_informacyjnego\\_wojewodztwa\\_slaskiego\\_do\\_roku\\_2015](http://www.e-slask.pl/article/strategia_rozwoju_spoleczenstwa_informacyjnego_wojewodztwa_slaskiego_do_roku_2015), (accessed: 12th June 2016).
- [47] J. Collis and R. Hussey, *Business research. A practical guide for undergraduate and postgraduate students*. New York: Palgrave Macmillan, 2003.
- [48] D. Gilliland and V. Melfi, “A note on confidence interval estimation and margin of error,” *Journal of Statistics Education*, vol. 18, no 1, pp. 1–8 (2010). [www.amstat.org/publications/jse/v18n1/gilliland.pdf](http://www.amstat.org/publications/jse/v18n1/gilliland.pdf) (accessed: 19th April 2019).
- [49] R.J. Thornton and J.A. Thornton, “Erring on the margin of error,” *Southern Economic Journal*, vol. 71, no. 1, pp. 130–135, 2004. <https://www.jstor.org/stable/4135315>.
- [50] E. Ziemia, Eds. *Czynniki sukcesu i poziom wykorzystania technologii informacyjno-komunikacyjnych w Polsce [Success factors for and level of ICT adoption in Poland]*. Warsaw: CeDeWu, 2015.

# 1<sup>st</sup> Special Session on Data Science in Health

**T**HE Special Session on Data Science in Health is a forum on all forms of data analysis, health economics, information systems and data based health service research, focusing mainly on the interaction of those four fields. Here, data-driven solutions can be generated by understanding complex real-world health related problems, critical thinking and analytics to derive knowledge from (big) data. The past years have shown a forthcoming interest on innovative data technology. Already now we can see how immense amounts of data and rapidly increasing, inexpensive computing power will lead the world to base its decisions more and more on data. We therefore have to work together. We need the knowledge of researchers from different fields applying diverse perspectives and using different methodological directions to find a way to grasp and fully understand the power and opportunities of big data in health.

This special session is a joint track by WIG2, the Scientific Institute for health economics and health service research, and the Information Systems Institute of Leipzig University.

## TOPICS

We embrace a rich array of issues on data science in health and offer a platform for research from diverse methodological directions, including quantitative empirical research as well as qualitative contributions. We welcome research from a medical, technological, economic, political and societal perspective.

The topics of interest therefore include but are not limited to:

- Data analysis in health
- Health Data management
- Health economics
- Data based health service research

- Integrating data in integrated care
- AI in integrated care
- Spatial health economics
- Structural equation modelling in medical research
- Risk adjustment and Predictive modelling

## EVENT CHAIRS

- **Franczyk, Bogdan**, University of Leipzig, Germany
- **Häckl, Dennis**, WIG2 Institute for health economics and health service research, Leipzig, Germany

## PROGRAM COMMITTEE

- **Alpkoçak, Adil**, Dokuz Eylul University
- **Dey, Nilanjan**, Techno India College of Technology, India
- **Kossack, Nils**, Head Mathematics and Statistics, WIG2 Institute for Health Economics and Health Service Research
- **Kozak, Karol**, Fraunhofer and Uniklinikum Dresden, Germany
- **Militzer-Horstmann, Carsta**, WIG2 Institute for Health Economics and Health Service Research, Information Systems Institute of the University of Leipzig, Germany
- **Popowski, Piotr**, Medical University of Gdańsk, Poland
- **Sachdeva, Shelly**, National Institute of Technology Delhi
- **Wasielewska-Michniewska, Katarzyna**, Systems Research Institute of the Polish Academy of Sciences, Poland
- **Wende, Danny**, WIG2 Institute for Health Economics and Health Service Research And Technical University Dresden



# Medical data exploration based on the heterogeneous data sources aggregation system

Andrzej Opaliński\*, Krzysztof Regulski\*, Barbara Mrzygłód\*, Mirosław Głowacki\*<sup>†</sup>,  
Aleksander Kania<sup>‡</sup>, Paweł Nastałek<sup>‡</sup>, Natalia Celejewska-Wójcik<sup>‡</sup>, Grażyna Bochenek<sup>‡</sup> and Krzysztof Śladek<sup>‡</sup>

\*AGH University of Science and Technology,  
al.A.Mickiewicza 30, 30-059, Krakow, Poland

<sup>†</sup>The Jan Kochanowski University, ul.Zeromskiego 5, 25-001, Kielce, Poland

<sup>‡</sup>II Chair of Internal Medicine, Faculty of Medicine, Jagiellonian University Medical College,  
Skawinska 8, 31-066, Krakow, Poland

**Abstract**—The paper presents the implementation and use of the IT system implemented in the Department of Pulmonology of The University Hospital in Cracow. The system integrates data from heterogeneous sources of therapy, diagnosis and medical test results of patients with Obstructive Sleep Apnea (OSA). The article presents the main architectural assumptions of the system, as well as an example of data mining analyzes based on the data served by the system. The example of the research aims to present the possibilities offered by the integration of clinical data in telemedicine and the diagnosis of patients with sleep disordered breathing that may lead to certain comorbidities and premature death.

## I. INTRODUCTION

**O**BSTRUCTIVE Sleep Apnea (OSA) is a widespread sleep disorder. It is estimated that the syndrome is present in approximately 5% of the general human population [1]. It is characterized by obstruction of the upper airway despite ongoing breathing efforts that lead to intermittent hypoxia and awakenings. The typical symptoms of OSA are loud snoring with pauses in breathing and daytime sleepiness. If untreated, OSA can lead to a number of severe medical conditions, mainly cardiovascular complications [2].

Polysomnography is the gold standard in the diagnosis of OSA [1]. Continuous positive airway pressure (CPAP) is the gold standard in OSA treatment [14], [9], [20]. Diagnostic process as well as the therapy requires access to information from many sources. Both patient history and clinical examination as well as polysomnography (PSG) results are considered. Obtained data has a very diverse form and is generated from many sources and by various devices (physician, PSG result, therapy devices etc.). So far, the data was collected in various places — paper documentation, patient registration system, PSG service system and SD cards of CPAP devices. In the present paper we demonstrate unique solution among another polysomnography software, assembling various data in the one system.

The work was realized as a part of fundamental research financed by the Ministry of Science and Higher Education, grant no. 16.16.110.663. The research was co-financed with National Scientific Leading Center project funds.  
This research was supported in part by PLGrid Infrastructure

The implemented system allows for detail research and data analysis leading to improvement of diagnostic quality and shortening its time by gathering all required data in one system [15]. Clinical data in connection with PSG results and CPAP recordings are input to the analysis of multidimensional and multicriteria links between individual indicators and other tests (i.e. blood lipids, arterial blood gases, glucose, creatinine etc.). It is expected that the development of research based on these indicators will enable the creation of a metamodel of mechanisms operating in this area in the future. At the moment, the aim of the research was to identify in which situations the basic diagnostic criteria (individual indicators mentioned above) fail.

For the initial cardiovascular risk assessment, the SCORE cardiovascular risk algorithm (Systematic COonary Risk Evaluation) is frequently used. The main risk factors for cardiovascular complications in the studied OSA cohort collected in the database include: (1) BMI (body mass index), (2) blood cholesterol, (3) systolic blood pressure, (4) package years — a factor calculated as a combination of years of smoking of cigarettes and the number of pieces smoked per day, (5) gender, (6) age.

Selected PSG parameters include: (1) AHI (Apnea Hypopnea Index) — number of apneas and hyponeas per hour of sleep), the most important OSA indicator used to determine its severity, (2) ODI (Oxygen Desaturation Index) — number of hemoglobin oxygen saturation falls per hour of sleep) demonstrate the level of sleep hypoxia.

AHI is the basic and the most common objectively used index to stage OSA severity. Judging the severity of disease we should take into account other tools including subjective doctor's opinion. This personal impression may have important value and can be measured by Clinical Global Impression Severity Scale (CGISS).

The above list shows a certain space of possible inaccuracies. Based on the medical history and physical examination, physician evaluating patient using this subjective scale decides of the urgency for PSG examination.

## II. RELATED WORKS

While the problem of the diagnosis and treatment of diseases related to OSA is a well known subject, the number of IT solutions supporting doctors in this field is relatively small. One of the works in this field mentioned in the literature is the system described by Passali et al. [16], which concerns the database of OSA patients undergoing upper airway surgery. Anthropometric data, results of scales diagnosing OSA occurrence, data from PSG tests, laryngological tests and laboratory tests were stored in the database. The collected data concerned the condition of patients before and after surgery, allowing their use as a source for methods supporting the automatic diagnosis of patients.

A separate extensive system that collects data on patients with OSA is the ESADA database [5], which integrates data from 22 medical units from all over Europe. The system stores data regarding the treatment of patients from the moment of diagnosis through the entire treatment process. Such a diverse range of patient groups allowed for a series of studies related to the detection of previously unknown dependencies, the causes of disease, including environmental and epidemiological conditions [19]. Based on data collected in the ESADA system, the relationship between OSA and problems related to hypertension [21], [22], kidney diseases [12] and diabetes [10] were also determined. These data also allowed to indicate the relationship between the use of different scales of diagnosis of OSA on the effectiveness and accuracy of the diagnosis process itself [2]. Due to the large number of data stored in the ESADA system, we can expect in the near future further publications of research results, developed on their basis.

When it comes to OSA data integration systems in individual countries, two such solutions have emerged in recent years. One of them was the Turkish TURKAPNE system (The Turkish Sleep Apnea Database), which began operation in 2017 and is to collect information about patients treated with OSA within the next 10 years. Another is the Danish NDOSA patients database [7], [8], which is assumed to collect data on the treatment of OSA-related conditions, in order to improve the quality of treatment in this field. Based on the publications data available in the literature, systems supporting treatment in the OSA field begin to appear in the medical market to facilitate and improve the process of diagnosis and treatment of patients. However, these are mostly databases themselves, without advanced diagnostic algorithms based on more extensive methods of data analysis

## III. SOLUTION CONCEPT AND SYSTEM ARCHITECTURE

In order to integrate all heterogeneous data sources, and ensure their consistency and security, an IT system was developed. It's main elements are presented in Fig. 1 and it consists of:

- A virtual central server maintained in the infrastructure of ACK Cyfronet ;

<http://www.cyfronet.krakow.pl/en/4421,main.html>

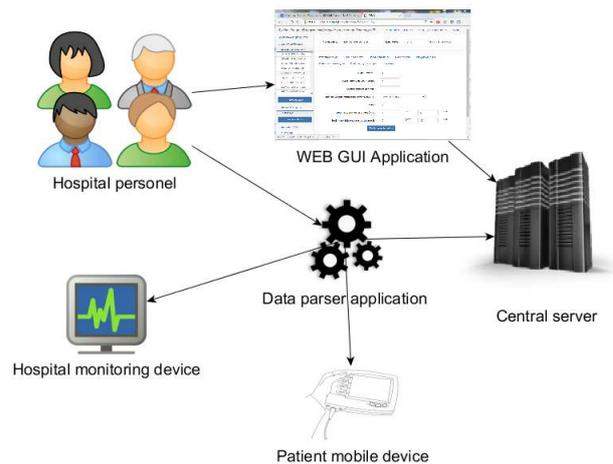


Fig. 1. Main system components

- A graphical user interface that provides access to the system for the hospital's medical staff (for entering data in non electronic format — diagnoses, lab results, etc.);
- Application that allows data to be exported to the system from two types of devices that monitor patients' sleep:
  - PSG device — advanced monitoring of patient's sleep during hospital stay;
  - CPAP devices — mobile patient devices that monitor patient's sleep during therapy outside the hospital.

The main application operating on the central server is based on the MVC (model-view-controller) architectural model developed with PlayFramework programming environment. The graphical user interface (GUI) of the application, which allows hospital staff to enter data and manage the system is presented in Fig. 2.

Datasets stored in the system are based on following:

- Clinical data — interview, physical examination, diagnosis, drugs;
- Diagnostic tests — PSG, CPAP, laboratory tests, spirometry;
- Medical recommendation — previous and planned treatment.

At the model layer, data is stored in a relational MySQL database, the structure of which is presented in Fig. 3. The business logic of the application is implemented in JAVA language on a virtualized CentOS operating system. The presentation layer is based on script templates in SCALA with HTML output code.

Based on the graphical user interface, hospital personnel provide the data in the system during the patient's visit to the hospital. In addition to standard questionnaires and medical diagnosis, data on further treatment as well as laboratory tests and lung tests — stored so far in external IT systems — are also provided.

Another key element of the system is the application that allows to extract data from devices that monitor the patient's

sleep. The current version of the system supports two types of devices. The first of these is the PSG device, which monitors the patient's sleep during his several-day visit to the hospital. The Sleepware3D application from Phillips Respironics is used to operate this device, and the data obtained from it is saved in RTF format. The second type of devices that monitors the patient's sleep during his stay outside the hospital are mobile versions of devices — CPAP, which generate reports in PDF format. In order to import selected reports from CPAP and PSG devices into the system, hospital staff run manually a dedicated application. The application developed within the system allows (using dedicated parsers algorithms) to extract data from device reports and export them to a central server that integrates them with other data related to a specific patient. PDFParser and RTFEditoKit JAVA libraries were used to extract data from documents in PDF and RTF formats as well as dedicated templates for extracting relevant data from these documents. Due to the fact that data import is performed for an individual patient, there is no problem of overloading the system during the import process.

Such integrated data, acquired from many heterogeneous sources, previously stored within various information systems and in paper form, are integrated into one universal data model (Fig. 3) within the presented system and made available for processing for advanced data processing and analysis methods.

#### IV. DATA EXPLORATION RESULTS

In the Introduction section, a research problem was initially drawn up. Data analysis presented in this research was performed to show the huge potential for integrated data collection in the IT system for their use in clinical practice. All computations was carried out on an integrated patient database, with the use of STATISTICA software. At first, the relationships between some patients' clinical data and their characteristics were investigated. The analysis showed that some dependencies exist. It should be stated here that particular variables could have different scales. Some of them

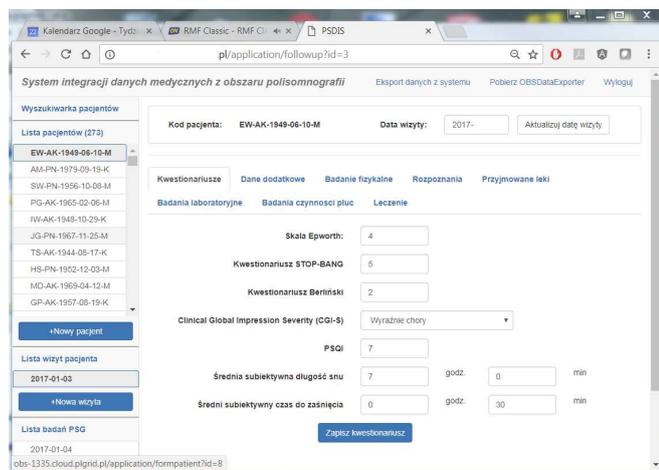


Fig. 2. WEB interface for system management

were qualitative (e.g. sex), other ordinal (e.g. measuring scales) and other quantitative (e.g. index of disorders). A series of tests on the dependence of parameters was performed — correlation test (Pearson's coefficient), t tests as well as chi square tests. Some of the dependencies can be presented in the charts Fig. 4.

From the charts of means in groups (Fig. 4) we can find the relationship between gender and the number of package-years and AHI. The AHI variable includes three classes that express the severity of the disease, where AHI = 3 means a severe OSA. We can observe a pattern that men with severe OSA smoke more. The question is that AHI depends on the amount of cigarettes smoked. The third graph shows that smoking correlates with AHI. While it cannot be demonstrated that smoking causes OSA, it can be assumed that perhaps smoking aggravates the severity of the disease. Such conclusions could be drawn from the tested sample it is certainly a pattern that is worth further research. Does severe OSA predispose to heavy nicotine dependence?

We can study the effect of risk factors on various OSA indices to find differences in their diagnostic strength. (Fig. 5). It can be seen that while age and BMI clearly affect the SCORE value, their impact on AHI and ODI (direct OSA indices) is small. This may negatively affect SCORE diagnostic capabilities, as evidenced by subsequent analyzes.

##### A. False Negatives Recognition

Looking for naturally occurring data structures, groups of patients with similar indicators and clinical characteristics, the method of clustering — unsupervised learning — k-means was used. Clustering is the most extensive group of machine learning methods called "unsupervised". Among the many known algorithms (EM algorithm, fuzzy c-means, Kohonen's Neural Networks, etc.) one of the oldest and most popular tools for the development of other tools is the k-means algorithm [11]. Clustering deals with searching for a structure in a set of unidentified data. It is the process of organizing objects into groups whose elements are in some way similar to each other [4]. In terms of computations, the algorithm is reduced to two-criteria optimization, where the distance between cluster objects is minimized, and the distance between clusters is maximized [6], [18].

The results are presented in Table I. The tests were successful, we managed to determine the concentration of patients with similar characteristics with a small error of validation. The distance between clusters was calculated by the Euclidian metrics, while the means and the most frequent values for descriptive variables are presented in Table I.

The analysis shows that 5 clusters can be distinguished from the patients (the number of clusters was set with cross-validation method): cluster 1, 3 and 4 these are cases of elevated AHI — which means that they group patients suffering from severe OSA. In contrast to clusters 2 and 5, which focus patients with mild form of the disease. It can be demonstrated using Table I that cluster 1 and 4 are: women with severe OSA and men with severe OSA — their CGIS (subjective

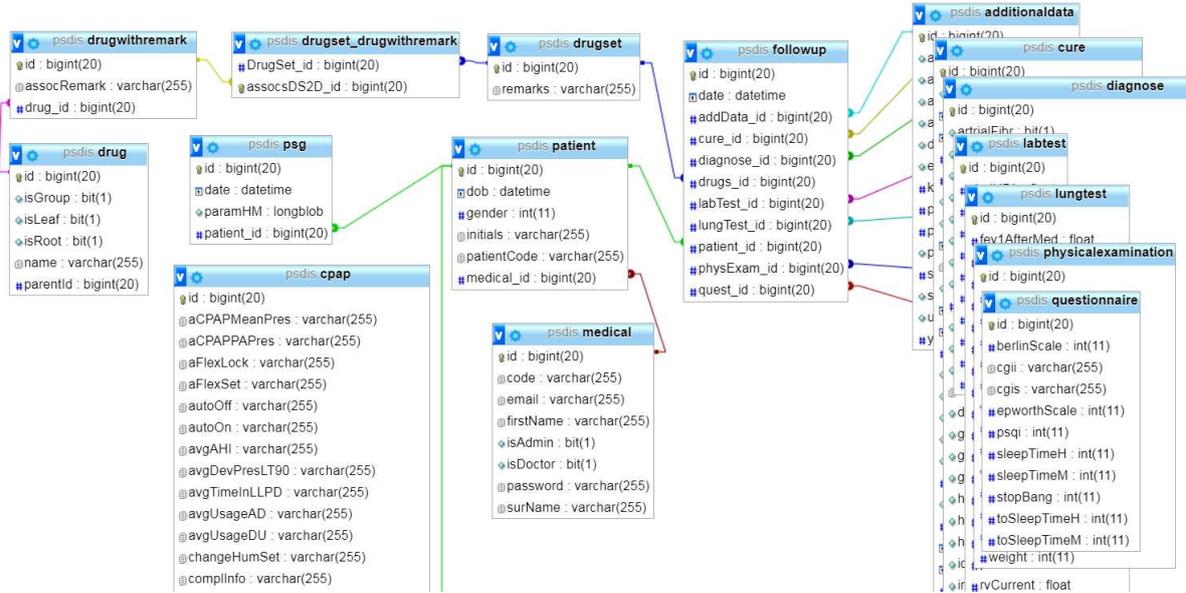


Fig. 3. Database structure

assessment of the doctor) and SCORE are high. Similarly, cluster 2 and 5 are the men and women with the lowest AHI — there CGIS and SCORE are low.

However, the analysis has identified yet another cluster — number 3. It is characteristic due to the high class of AHI (this class means that in the PSG examination the patient had the Apnea-Hypopnea Index >30), and at the same time low SCORE and CGIS. This means that they are possibly false negatives cases — patients with risk of having undiagnosed OSA (if the decision on referral for PSG examination should be made on the basis of machine learning methods).

Statisticians call this situation False Negatives. In statistical hypothesis testing false negatives are type II errors, where a negative result corresponds to not rejecting the null hypothesis. We can call it an underdiagnoses error. The conducted research

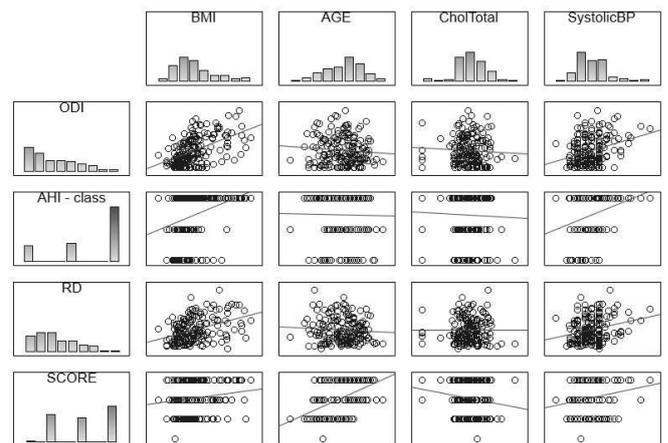


Fig. 5. The impact of risk factors on various OSA indicators

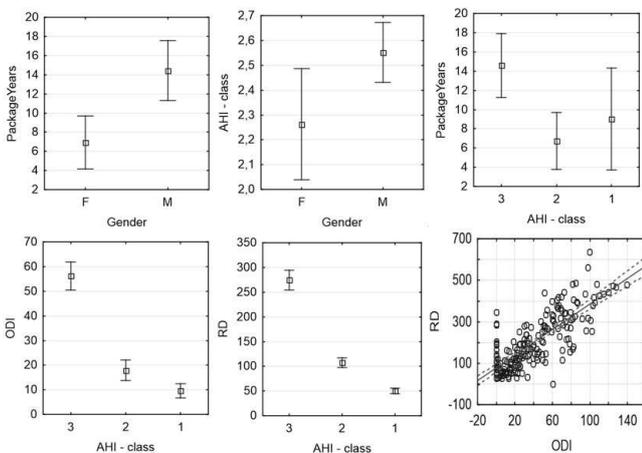


Fig. 4. Selected relationships between patients' characteristics

has shown that it is possible to select risk groups — patients whose diagnosis may be subject to the error false negatives, especially if the diagnosis would be carried out without PSG examination.

Is it possible to use this knowledge in future diagnoses? How machines can predict that a patient may belong to the FalseNegatives group and protect him against a mistaken underdiagnoses?

Cluster 3 is men who smoke less than other patients, who is also the youngest group among the respondents. They have high BMI, elevated blood pressure and cholesterol at the same time.

### B. False Negatives Prediction

In order to create a classification model, the algorithm for creating CART classification trees was used. Decision trees

TABLE I  
K-MEANS CLUSTER CHARACTERISTICS

Cluster	Gender	Age	CGIS	Package Years	BMI	Systolic BP	SCORE	Chol Total	AHI dominant	No of cases	(%)
1	F	58,4	3,6	8,5	40,0	140,5	3,15	4,29	3	32	17,0
2	F	59,3	3,0	4,2	30,1	132,2	2,80	5,27	1	25	13,2
3	M	43,1	3,0	9,2	34,3	141,9	2,20	4,90	3	29	15,4
4	M	62,8	3,7	19,5	34,8	142,4	3,78	4,41	3	64	34,0
5	M	56,7	2,8	10,3	28,4	133,2	2,71	4,46	1	38	20,2

are very popular among data mining tools [3]. Repeatedly described in the literature, they have found countless examples of applications [17], [21]. Their usefulness has been determined by the following characteristics: (1) easy to interpret by a human; (2) simple representation of complex relationships occurring in data sets; (3) no assumptions on the variability of input and output parameters, and (4) no assumptions on the probability distribution of variables; (5) the ability to operate on incomplete and noisy sets [13].

The algorithm evaluates the discriminative power of variables and chooses for division successively those that provide better separation of objects between classes. Then the split point is selected. The lowest Gini Index is chosen as the best dividing point. This process is repeated until a satisfactory tree is obtained (based on the leaf size or total classification error).

Fig. 6 presents the matrix of errors for the inducted CART classification tree. Using the tree we can predict that the patient belongs to the group of patients at risk of underdiagnoses mistake in 82.7% of cases. We reduce the risk of misdiagnosis by as much as a percentage if we consider the indications from the analysis.

Because false negatives cases mainly concern men, in graph in Fig. 7 only a fragment of the tree about men is presented.

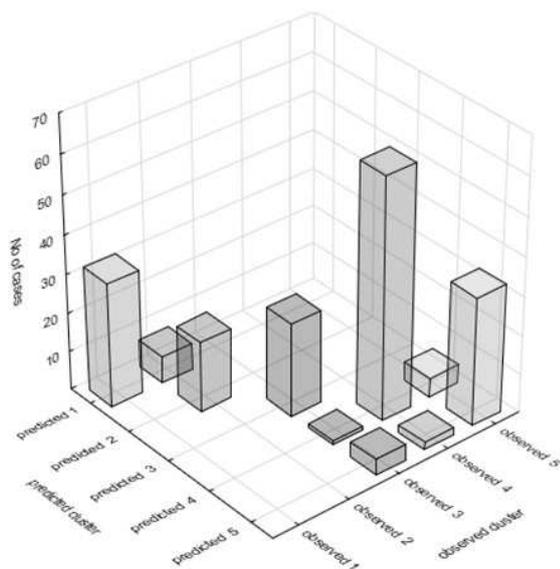


Fig. 6. The misclassification matrix for the False Negatives classification tree

Only in this area it is possible to predict the cluster 3.

Reading the rules induced by the tree we can notice: If someone is a man under 54 years of age, with a SCORE index below 3, it belongs to cluster 3 (FN) with 65% of certainty. If we additionally take into account the results of spirometry, including Tiffeneau index after bronchodilator below 86, the probability increases to 73.6%. Taking into account laboratory tests, it is worth paying attention to the content of C-reactive protein. If it is above 2 we are sure that it is FN (cluster 3), otherwise we check if the patient is younger than 43 years, then it certainly belongs to cluster 3. The remaining patients belong to cluster 5 — means true positives — patients with mild OSA.

When interpreting the results, it should be taken into account that the obtained results are susceptible to an error resulting from a small number of data stored so far in the system, and thus a relatively small number of variables that can be used in the analyzes. In the future, successive algorithms (decision trees and clustering) should operate on disjoint sets of attributes to avoid problem of endogeneity.

V. SUMMARY

This paper presents the design, implementation and preliminary results of a prototype system that enables to improve the diagnosis and possible decision-making about treatment of patients with sleep disorders. The main advantages of the system are aggregation of data from various sources (diagnostics, lab-tests, medical history) and its integration with devices for OSA diagnosis (PSG) and treatment (CPAP). It can significantly improve the work of the hospital staff and facilitate their access to previously distributed patient data.

Besides, the examples of data analyzes that allow searching for dependencies between clinical tests and diagnosis have been shown. Data mining analyzes allowed to find the characteristics of a group of patients for whom the risk of underdiagnoses was the highest. However, from a medical point of view, all presented results and obtained dependencies should be treated with extreme caution, because a small number of randomly selected parameters does not necessarily reflect reality. At this stage, the paper has only statistical and IT value, and its purpose was to show the huge potential for integrated data collection in the IT system for their use in clinical practice. The intention of the presented system is to improve the quality of diagnosis and treatment of patients affected by OSA and it seems that this goal is achieved after implementation of the system into daily clinical practice.

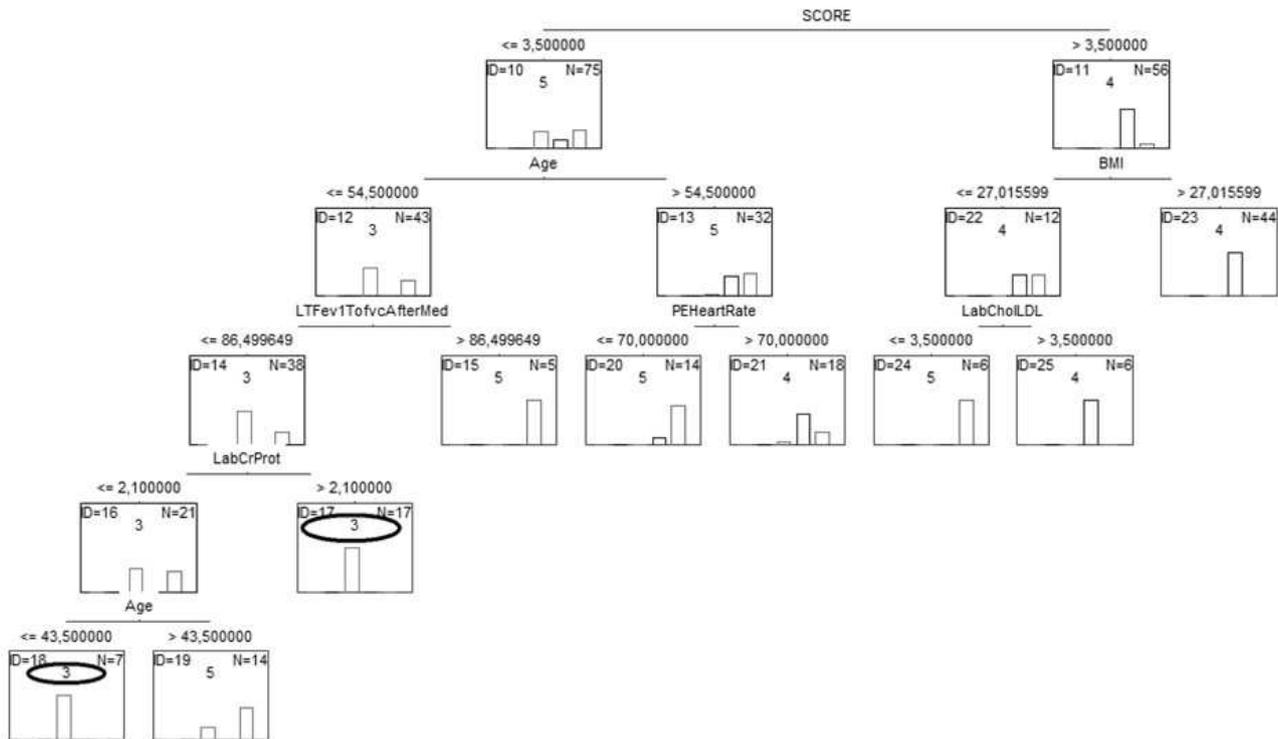


Fig. 7. Fragment of the classification tree

## REFERENCES

- [1] H.Y. Chiu, P.Y. Chen, L.P. Chuang, N.H. Chen, Y.K. Tu, Y.J. Hsieh, Y.C. Wang, and C. Guilleminault. Diagnostics accuracy of the berlin questionnaire, stop-bang, stop, and epworth sleepiness scale in detecting obstructive sleep apnea: A bivariate meta-analysis. *Sleep Medicine Reviews*, 36:57–70, 2016.
- [2] P. Escourrou, L. Grote, T. Penzel, W. T. McNicholas, J. Verbraecken, R. Tkacova, and F. Barbé. The diagnostic method has a strong influence on classification of obstructive sleep apnea. *Journal of sleep research*, 24(6):730–738, 2015.
- [3] A. Glowacz. Acoustic based fault diagnosis of three-phase induction motor. *Applied Acoustics*, 137:82–89, 2018.
- [4] J. A. Hartigan and M. A. Wong. Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society*, 28(1):100–108, 1979.
- [5] J. Hedner, L. Grote, M. Bonsignore, W. McNicholas, P. Lavie, G. Parati, P. Sliwinski, F. Barbé, W. De Backer, P. Escourrou, I. Fietze, J. A. Kvanne, C. Lombardi, O. Marrone, J. F. Masa, J. M. Montserrat, T. Penzel, M. Pretl, R. Riha, D. Rodenstein, T. Saaresranta, R. Schulz, R. Tkacova, G. Varoneckas, A. Vitols, H. Vrints, and J. Zielinski. The european sleep apnoea database (esada): report from 22 european sleep laboratories. *Eur Respir J*, 38:635–42, 2011.
- [6] A. K. Jain. Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*, 31(8):651–666, 2010.
- [7] P. Jennum, R. Ibsen, and J. Kjellberg. Morbidity prior to a diagnosis of sleep-disordered breathing: a controlled national study. *J Clin Sleep Med*, 9(2):103–108, 2013.
- [8] P. J. Jennum, P. Larsen, C. Cerqueira, T. Schmidt, and P. Tønnesen. The danish national database for obstructive sleep apnea. *Clinical Epidemiology*, 8:573–576, 2016.
- [9] Riha R.L. Jennum, P. Epidemiology of sleep apnoea/hypopnoea syndrome and sleep-disordered breathing. *Eur Respir J*, 33:907–14, 2009.
- [10] B. D. Kent, L. Grote, M. R. Bonsignore, T. Saaresranta, J. Verbraecken, and P. Lévy. Sleep apnoea severity independently predicts glycaemic health in nondiabetic subjects: the esada study. *European Respiratory Journal*, 44(1):130–139, 2014.
- [11] J. B. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Symposium on Math*, pages 281–297, Statistics, and Probability, Berkeley, CA, 1967. University of California Press.
- [12] O. Marrone, S. Battaglia, P. Steiropoulos, O.K. Basoglu, J.A. Kvanne, S. Ryan, J.L. Pepin, J. Verbraecken, L. Grote, J. Hedner, and M.R. Bonsignore. Chronic kidney disease in european patients with obstructive sleep apnea: the esada cohort study. *J Sleep Res*, 25:739–45, 2016.
- [13] E. Nawarecki, S. Kluska-Nawarecka, and K. Regulski. *Multi-aspect character of the man-computer relationship in a diagnostic-advisory system*, pages 85–102. Springer-Verlag, 2012.
- [14] American Academy of Sleep Medicine. *International classification of sleep disorders: diagnostic and coding manual*. Amer Academy of Sleep Medicine, Westchester, Illinois, USA, 2005.
- [15] A. Opaliński, P. Nastalek, B. Mrzygłód, N. Celejewska-Wójcik, M. Głowacki, G. Bochenek, K. Regulski, K. Śladek, and A. Kania. The system for integration of heterogeneous data sources in the domain of obstructive sleep apnea. In *Economic Advance In Behavioral and Sociocultural Computing (B. E. S. C. ) eds Economic*, editors, *Proc.Conf. 4th International Conference on Behavioral*, pages 1–6. Demazeau Y, 2017.
- [16] D. Passali, G. Caruso, L.C. Arigliano, F.M. Passali, and L. Bellussi. Arigliano lc, passali fm, bellussi i. database application for patients with obstructive sleep apnoea syndrome. *Acta Otorhinolaryngol Ital*, 32:252–255, 2012.
- [17] J. R. Quinlan. *Induction on Decision Trees, Machine Learning*. Kluwer Academic Publishers, Boston, 1986.
- [18] K. Regulski, D. Wilk-Końcodziejczyk, and G. Gumienny. Comparative analysis of the properties of the nodular cast iron with carbides and the austempered ductile iron with use of the machine learning and the support vector machine. *The International Journal of Advanced Manufacturing Technology*, 87(1):1077–1093, 2016.
- [19] T. Saaresranta, J. Hedner, M. R. Bonsignore, R. L. Riha, W. T. McNicholas, T. Penzel, U. Anttalainen, J. A. Kvanne, M. Pretl, P. Sliwinski, and J. Verbraecken. Clinical phenotypes and comorbidity in european sleep apnoea patients. *PLoS One*, 11(10), 2016.
- [20] White D.P. Amin R. et al. Somers, V.K. Sleep apnea and cardiovascular disease: an american heart association/american college of cardiology foundation scientific statement from the american heart association

- council for high blood pressure research professional education committee, council on clinical cardiology, stroke council, and council on cardiovascular nursing. in collaboration with the national heart, lung, and blood institute national center on sleep disorders research (national institutes of health). *Circulation*, 118:1080–111, 2008.
- [21] Y. Song and Y. Lu. Decision tree methods: applications for classification and prediction. *Shanghai Arch Psychiatry*, 27(2):130–135, 2015.
- [22] R. Tkacova, W. T. McNicholas, M. Javorsky, I. Fietze, P. Sliwinski, G. Parati, L. Grote, and J. Hedner. Nocturnal intermittent hypoxia predicts prevalent hypertension in the european sleep apnoea database cohort study. *Eur Respir J*, 44:931–41, 2014.



# Mapping of Dental Care in the Czech Republic: Case Study of Graduates Distribution in Practice

Matěj Karolyi

Faculty of Informatics, Masaryk University, Botanická 68a, Brno, 602 00, Czech Republic  
and  
Institute of Health Information and Statistics of the Czech Republic  
Palackého nám. 4, 128 01, Praha 2, Czech Republic  
Email: karolyi@iba.muni.cz

Jakub Ščavnický,  
Martin Komenda

Institute of Health Information and Statistics of the Czech Republic and Institute of Biostatistics and Analyses, Faculty of Medicine, Masaryk University - joint workplace,  
Palackého nám. 4, 128 01, Praha 2, Czech Republic  
Email: {scavnicky, komenda}@iba.muni.cz

Jan Bud'a, Tereza Jurková,  
Monika Mazalová

Faculty of Science, Masaryk University, Kotlářská 2, Brno, 625 00, Czech Republic  
Email: {451441, 451627, 451455}@muni.cz

**Abstract**—Online registers contain a large amount of data about healthcare providers in the Czech Republic. Information is available to all citizens and can be useful to patients, governmental organisations or employers. Based on these data, we are able to create a high-quality snapshot of the current state of healthcare providers. Interconnecting data from more data sources together is an interesting task, and accomplishing it enables us to ask more complex questions. This paper focuses on answering several questions about dentists in our country. A dataset from one online database was created, using automated data mining methods and a subsequent analysis. Results are presented via an online tool, which was provided to owners of the data. They reviewed our results and decided to use our findings for the presentation to the Czech government and subsequent negotiation processes. Our paper describes used methods, shows some results and outlines possibilities for further work.

## I. INTRODUCTION

Dental care coordinated by the Czech Dental Chamber is an integral part of the healthcare system in the Czech Republic. Various dental services are provided by more than 8,000 dentists working at university clinics or municipal health centres as well as by private dentists and dental laboratories<sup>1</sup>. More than half of these dental practices are based in big cities, including the Capital of Prague. Universities provide dental study programmes fully in accordance with standards of the European Union. Dentistry curricula are completely separated and independent from general medicine study programmes. In general, dentistry programmes cover all basic requirements to practice dentistry in terms of prevention, diagnosis, treatment, medical opinion and monitoring. Students achieve knowledge and skills of all the activities and interventions as regards prevention, diagnosis and treatment of anomalies and diseases of the teeth, gums, jaws, and surrounding tissues<sup>2</sup>. From the government and higher education perspective, many online data sources describing the field of dentistry on both national

<sup>1</sup> <https://www.dent.cz/>

and regional levels are available. This information can provide interesting inputs for further analyses which explore new relations and patterns between graduates and practices.

### A. Portals presenting national dental care professionals

The network of healthcare providers in the Czech Republic is complex and well-described. The Ministry of Health of the Czech Republic and its departments provide searchable databases of all registered providers. These databases, which are available to all citizens, were included in a complex national review [1] of publicly available web portals together with others mainstream websites providing information from healthcare and medicine. The major online databases of individual healthcare providers and organisations are listed in Table I.

Table I  
Major Czech databases of healthcare providers

Name of the portal	Reference
Czech Medical Chamber	<a href="http://www.lkcr.cz">www.lkcr.cz</a>
Czech Dental Chamber	<a href="http://www.dent.cz">www.dent.cz</a>
National Register of Healthcare Providers	<a href="http://nrpzs.uzis.cz">nrpzs.uzis.cz</a>
Open Data of the Ministry of Health of the Czech Republic	<a href="http://opendata.mzcr.cz">opendata.mzcr.cz</a>
Open Data of the State Institute for Drug Control	<a href="http://opendata.sukl.cz">opendata.sukl.cz</a>
Portal of Advisory Bodies, Working Groups and Expert Committees of the Ministry of Health of the Czech Republic	<a href="http://ppo.mzcr.cz">ppo.mzcr.cz</a>
Portal for Patients and Patient Organisations	<a href="http://pacientskeorganizace.mzcr.cz">pacientskeorganizace.mzcr.cz</a>
ZnamyLekar	<a href="http://www.znamylekar.cz">www.znamylekar.cz</a>

In this paper, we focus on the second of the above-mentioned databases, which is guaranteed by the Czech

<sup>2</sup> <https://www.muni.cz/en>

Dental Chamber. These data fit most conveniently to our further investigation because they contain information only about dentists, not about other health professionals. The obtained dataset from the publicly available database is therefore as relevant as it can be for further examination.

### *B. Motivation and exploratory questions*

On the one hand, a lot of data describing dental care in the Czech Republic are freely available. In theory, there are no limitations and borders to mine and to process those data. On the other hand, a huge amount of records on individual dental care providers make a global overview and orientation in the particular domain of dentistry on the national level quite complicated and unclear. Moreover, the manual process of data extraction and local database construction is very time-consuming.

This paper aims to find an effective way of extracting data automatically from freely accessible online sources using a machine-based – instead of a human-based – approach. With respect to our other research activities [2]–[4], we decided to explore the domain of dental care from two different perspectives: (i) Czech higher education institutions, which guarantee various dental medicine study programmes, (ii) real distribution of dental professionals in everyday clinical practice. The process of mapping of dental care in the Czech Republic in terms of graduates' distribution across the country was a challenge from the very beginning. A student project devoted to this particular topic was solved at the Faculty of Science of the Masaryk University. Based on data from the Czech Dental Chamber portal, a pilot automated mapping between graduates and dental professionals was done. Finally, a web-based application presenting the achieved results in the form of an interactive visualisation has been designed, developed and implemented.

## II. METHODS

The preparation of a final output (i.e. the online visualisation tool in this case) had several stages: obtaining the dataset, data preparation for further analysis, development of multiple interactive views and a final evaluation. All activities were carried out by a team of three students under the supervision of mentors from the Web Design Department<sup>3</sup> of the Institute of Biostatistics and Analyses at the Faculty of Medicine of the Masaryk University (IBA FM MU). During the process of data mining, we followed the standardised and proven methodology called the cross-industry standard process for data mining (CRISP-DM). It helped us to avoid the common mistakes and to work efficiently as a team [5]. We have distributed our activities in this case study, too. The next sections describe our steps in the context of CRISP-DM.

### *A. Business and data understanding*

The web portal [dent.cz](http://dent.cz) provides information for members of the Czech Dental Chamber (CDC) as well as for the general public.

The portal consists of the following sections:

- list of dentists,
- education – a calendar of events, recommended literature and other study materials,
- LKS journal – information about their periodical,
- news – current events and news in the dentistry,
- about us – general information about CDC,
- for members – accessible only to CDC members,
- contacts – contact to the CDC office.

In particular, we were interested in the very first item, i.e. the list of dentists, for further investigation.

Records on particular dentists are available through records on individual healthcare facilities, and each dentist can be registered at none, one or more of these facilities. All records have a clearly defined common structure consisting of the dentist's name, information about his/her workplace, education and regional dental chamber. The section about healthcare facility where the dentist works is the key part of the record. It consists of the name of the healthcare facility, its address and contact. Three ways of filling this section are distinguished. In the first case, the dentist works only in one facility. In the second case, the dentist works in more than one facility. Finally, no healthcare facility is mentioned. Information about the dentist's workplace(s) is supplemented by a map.

The education section was another important part of this study. As was the case of healthcare facilities, it was filled in three different ways (one university, more than one university, no university mentioned).

### *B. Data preparation*

Data preparation consisted of two main steps. The first one involved web mining methods and the insertion of gathered pieces of information about dentists in a structured form into the database. The subsequent phase focused on data cleaning, which meant extracting useful analytical information by regular expressions from HTML codes into a new table in the database. All steps in this section were created in the Python programming environment using libraries that are described in Section 2.2.1. In the following text, more detailed information about the algorithm we have designed will be provided.

Web mining is generally called crawling [6] because the algorithm goes gradually through the web portal hierarchy. The crawling algorithm has two functionally different parts. The first part consists of many functions which work with URL links. The function for extracting all URLs from a specific web page is the most important segment of the code, in which several key conditions are defined. For instance, we had to select only unique URLs from the list of all URLs on the page, and we needed to ensure the algorithm would be terminated when the 'offset' was detected in the crawled URL. After we obtained the final list of URLs, we

<sup>3</sup> <http://www.iba.muni.cz/index-en.php?pg=contract-research--web-design>

created the function for scraping a specific URL. This is represented by extracting the HTML code, which includes all the information about dentists in the free-form text, from the web page.

In the second part of the proposed crawling algorithm, the obtained information was inserted into a newly created database. Firstly, the connection to a SQLite Database Server was created, then a table for crawled URLs was created and the first record was added to the database. Secondly, all URLs from the first web page were inserted into the database table and the table was updated with the HTML code of this page. This process was iterative until the HTML codes of all records in the table were filled.

The proposed methods of the extraction algorithm mentioned in the introduction of this section have two parts as well. In the first one, functions for extraction of information about dentists were created, using various regular expressions from the HTML codes. In this part of the code, names of schools had to be unified because there was an enormous inconsistency in foreign school names. Firstly, a new table for the extraction was created, then only the records about dentists (not about healthcare facilities) were selected from the primary table. Secondly, all extracted attributes were inserted into the database table at the same time. In this key step, the issue with more healthcare facilities per dentist was resolved by a uniform distribution of the Full-time equivalent (FTE) among the workplaces. For example, if a dentist worked in three workplaces, then the weight of each record about this dentist was 0.33. For further interactive data analyses and visualisations, it was crucial to extract the names of healthcare facilities as well, since each healthcare facility was defined with respect to its geographical location as a unique combination of the name of the relevant facility, its latitude and longitude. Subsequently, it was necessary to obtain further information about a given region and a district workplace using the postal code of that workplace, using data from the web portal <http://www.psc.cz>. This information was automatically extracted from the HTML code of this web portal using the above-mentioned method. The postal code of each workplace was primarily used to assign both the region and the district to each healthcare facility.

Using this procedure, however, we were not able to assign all regions and districts, so we subsequently decided to use the municipality where a given healthcare facility was located for search on the <http://www.psc.cz> portal. In this manner, the number of healthcare facilities with unknown regions and districts was significantly reduced. The proposed extraction algorithm was conducted with a SQL update of the table and thus the final version of the dataset for further data analysis was obtained.

### C. Modeling and Evaluation

Extracted and cleaned data saved in the SQLite Database Server was connected with the R programming language. Afterwards, SQL queries were executed over the database

using the R programming environment. These data aggregations were used for the creation of R Shiny application, especially for descriptive statistics and visualisations that were represented for example by textual descriptions donut charts or cartograms.

The application was independently evaluated twice in the work team: within the students' team and then in the mentors' team. The final output was also presented to other teams of the subject and their mentors. The whole auditorium had an opportunity to participate in the discussion. Subsequently, the application was presented to representatives of the CDC. They commented on factual accuracy and usability of the presented outputs. Finally, all remarks collected during the review process were incorporated into the application.

### D. Technological background and deployment

Various technologies, tools and packages were used during the deployment process. We are able to divide the technologies we used into two categories by their purpose within the whole project: (i) data retrieval group – tools and libraries which were used to obtain data from the web portal of the Czech Dental Chamber, (ii) data visualisation group – tools used in the final presented application for computing and rendering the user views with graphs and text information.

#### Data retrieval group

The technological group of data retrieval consisted of scraping [7], database operations, data cleaning and data parsing. Each of the related procedures and methods were performed using the Python 3 language. Packages like urllib, requests, BeautifulSoup and sqlite3. SQLite was used as the application's database layer.

#### Data visualisation group

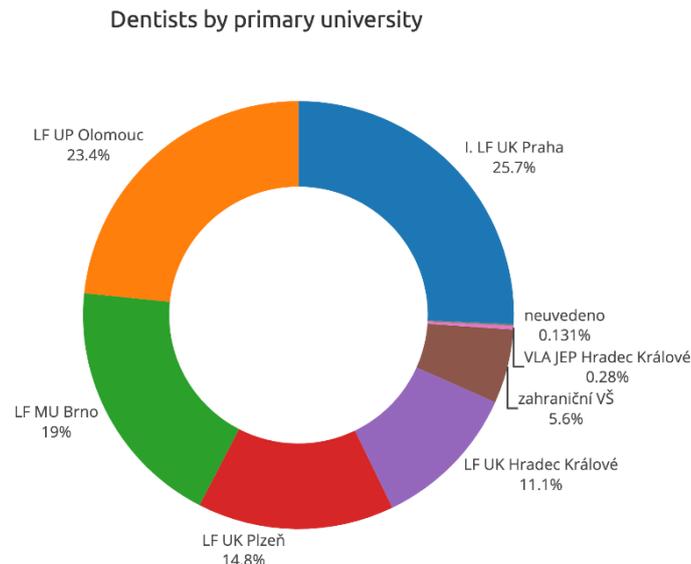
The technological group of data visualisation consisted of data aggregation, charts plotting and app deployment. Each of the related procedures and methods were performed in the R software environment. The RSQLite package was used for communication with the SQLite database engine in order to aggregate data effectively. Packages plotly and networkD3 were used to create interactive pie charts and a map of the Czech Republic. The whole application was built as an interactive R application using the packages Shiny and shinythemes. Furthermore, it was deployed on our Open CPU server. Therefore, our R Shiny applications allow real-time user interaction and data filtering in a simple online environment.

## III. RESULTS

We have created a publicly available tool which shows the state of distribution of dental care graduates all over the Czech Republic. The dataset was collected on 22 October 2018. The tool is available on a public URL<sup>4</sup> with the user interface translated to English.

<sup>4</sup> <https://vis.iba.muni.cz/apps/dent-en/>

## Number of dentists: 10726



Note: The chart does not include dentists who have studied multiple schools. Specifically, there are 20 cases.

Fig. 1 Percentage of dentists by universities at which they studied primarily

### A. Basic description and overview of studies

The application was created as an output of a student project. On the initial tab of the web application, the objective and basic information are mentioned.

The second tab deals with the education of dentists, precisely with their primary school in relation to their dental practice. Six faculties of medicine in the Czech Republic were distinguished:

- First Faculty of Medicine of the Charles University in Prague,
- Faculty of Medicine and Dentistry of the Palacký University in Olomouc,
- Faculty of Medicine of the Masaryk University in Brno,
- Faculty of Medicine in Plzeň of the Charles University,
- Faculty of Medicine in Hradec Králové of the Charles University,
- Jan Evangelista Purkyně Military Medical Academy in Hradec Králové.

All foreign universities were united into one category and another category was created by merging dentists with missing information on university at which they studied.

The total number of dentists registered on the website of the Czech Dental Chamber was 10,726. Twenty of them mentioned two different universities at which they had studied. It was not possible to identify which of these universities was the one at which they had studied primarily, therefore these dentists were not included in further analyses.

Fig. 1 shows the percentage of dental practitioners by universities at which they studied primarily. Most dentists

graduated from the First Faculty of Medicine of the Charles University in Prague (25.70%), the Faculty of Medicine and Dentistry of the Palacký University in Olomouc was the second most frequently mentioned one (23.40%), and the Faculty of Medicine of the Masaryk University in Brno was the third one (19.00%). The proportion of dentists who studied abroad was 5.60%. In fourteen cases, the dentist's education was unknown (0.13%).

### B. Dental offices in the Czech Republic

Dental offices in the Czech Republic are displayed on two tabs (the third one and the fourth one). The first of them describes only dental offices regardless of information on university graduates. About a fifth (21.00%) of dentists did not mention the healthcare facility in which they worked. More than two thirds (70.10%) of dentists worked in just one office and 8.90% of them worked in several offices. Therefore, a new variable was created – work time. Connection of work time to the FTE is described in more detail in chapter Data preparation. The third tab also displays maps containing information about numbers of healthcare facilities (6,579 in total) and work time of dentists by region. The cartogram allows user interactivity in the form of radio buttons. The user can choose from two options: the first of them shows the numbers of work times, whereas the second option displays the numbers of healthcare facilities by region. The first option can be seen in Fig. 2. This image clearly shows that the majority of dentists work in the capital (Prague) and its vicinity (Central Bohemian Region), followed by the South Moravian Region and the Moravian-Silesian Region. It is also obvious that the Karlovy Vary Region, has the lowest work time in the Czech Republic.

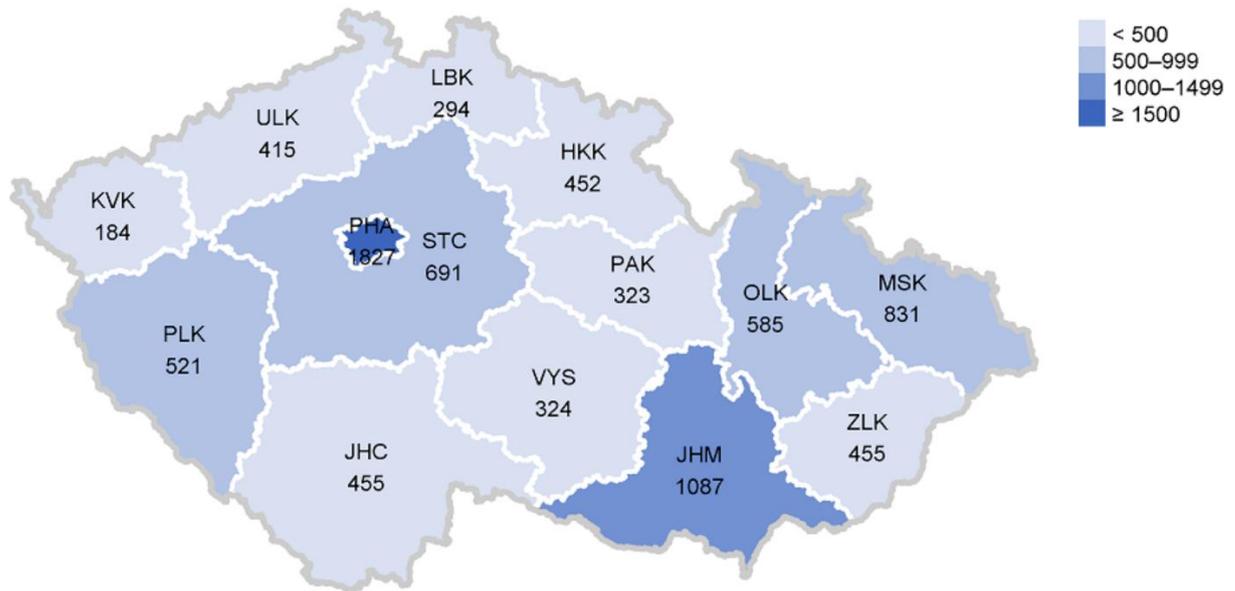


Fig. 2 Numbers of work time by region.

#### IV. DISCUSSION

This paper represents a student project solved under the supervision of senior mentors, where a proof of concept in online data crawling and scraping was carried out. Real data from a guaranteed online source on dental care were automatically mined and processed by a set of Python algorithms and stored in a relational database running on our own servers. The challenging task of complex mapping between data describing the distribution of dental professionals in practice was successfully solved. The final application is one from a group of similar projects [8], [9] being solved at the Institute of Biostatistics and Analyses at the Faculty of Medicine of the Masaryk University. It provides an original overview of data stored in the portal of the Czech Dental Chamber because it reveals hidden relations between graduates' and dental professionals' distribution across the Czech Republic.

The public R Shiny application and its outputs were subsequently consulted with representatives of the Czech Dental Chamber. Conclusions of the review were considered and implemented in the application. We believe that information obtained in this way will serve to increase the transparency of healthcare in the Czech Republic and will be an interesting source of knowledge for the entire community associated with the Czech Dental Chamber.

The proposed application is still open to changes and improvements. Updating the underlying dataset at different times would also be worth considering. It would then be possible to compare the evolution of migration over time and monitor the increment / decline of registered dentists across certain periods. In the future, it would be interesting to include demographic data from individual regions of the Czech

Republic in the analysis. It would then be possible to estimate the number of citizens in a certain region per one dental practitioner and whether a certain region is lacking this type of healthcare providers.

#### REFERENCES

- [1] M. Karolyi and M. Komenda, 'PŘEHLED ELEKTRONICKÝCH INFORMAČNÍCH ZDROJŮ VE ZDRAVOTNICTVÍ ČR', *MEDSOFT 2019*, p. 5.
- [2] L. Dušek, J. Mužík, M. Karolyi, M. Šalko, D. Malůšková, and M. Komenda, 'A Pilot Interactive Data Viewer for Cancer Screening', in *Environmental Software Systems. Computer Science for Environmental Protection: 12th IFIP WG 5.11 International Symposium, ISESS 2017, Zadar, Croatia, May 10-12, 2017, Proceedings 12*, 2017, pp. 173–183.
- [3] C. Vaitis et al., 'Standardization in medical education: review, collection and selection of standards to address', *MEFANET J.*, vol. 5, no. 1, pp. 28–39, Nov. 2017.
- [4] M. Komenda, M. Karolyi, C. Vaitis, D. Spachos, and L. Woodham, 'A Pilot Medical Curriculum Analysis and Visualization According to Medbiquitous Standards', in *2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS)*, 2017, pp. 144–149.
- [5] R. Wirth, 'CRISP-DM: Towards a standard process model for data mining', in *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*, 2000, pp. 29–39.
- [6] S. vanden Broucke and B. Baesens, *Practical Web Scraping for Data Science: Best Practices and Examples with Python*. Apress, 2018.
- [7] R. Mitchell, *Web scraping with Python: collecting data from the modern web*, First edition. Sebastopol, CA: O'Reilly Media, 2015.
- [8] L. Woodham, J. Ščavnický, M. Karolyi, and M. Komenda, 'Interactive presentation of evaluation data in training against medical errors', *Masarykova univerzita*, 2018. [Online]. Available: <https://www.muni.cz/vyzkum/publikace/1476359>. [Accessed: 30-Apr-2019].
- [9] M. Komenda, J. Ščavnický, P. Růžičková, M. Karolyi, P. Štourač, and D. Schwarz, 'Similarity Detection Between Virtual Patients and Medical Curriculum Using R', *Stud. Health Technol. Inform.*, vol. 255, pp. 222–226, 2018.



# Medical prescription classification: a NLP-based approach

Viincenza Carchiolo, Alessandro Longheu  
Universita di Catania  
Email: vincenza.carchiolo@unict.it  
alessandro.longheu@dieci.unict.it

Giuseppa Reitano, Luca Zagarella  
Previnet s.p.a. Treviso, Italy  
Previmedical s.p.a. Treviso, Italy  
Email: {giuseppa.reitano, luca.zagarella}@previnet.it

**Abstract**—The digitization of healthcare data has been consolidated in the last decade as a must to manage the vast amount of data generated by healthcare organizations. Carrying out this process effectively represents an enabling resource that will improve healthcare services provision, as well as on-the-edge related applications, ranging from clinical text mining to predictive modelling, survival analysis, patient similarity, genetic data analysis and many others. The application presented in this work concerns the digitization of medical prescriptions, both to provide authorization for healthcare services or to grant reimbursement for medical expenses. The proposed system first extract text from scanned medical prescription, then Natural Language Processing and machine learning techniques provide effective classification exploiting embedded terms and categories about patient/doctor personal data, symptoms, pathology, diagnosis and suggested treatments. A REST ful Web Service is introduced, together with results of prescription classification over a set of 800K+ of diagnostic statements.

## I. INTRODUCTION

In recent years, there has been an amplified focus on the use of Artificial Intelligence (AI) in E-health. There are numerous examples that include AI approaches to analyze unstructured data such as photos, videos, physician notes to enable clinical decision making; or the use of intelligent interfaces to enhance patient engagement and compliance with treatment and predictive modelling to manage patient flow and hospital capacity/resource allocation.

Two main information sources play a relevant role in healthcare field, i.e. images and natural language. The use of Natural Language Processing (NLP) found several applications related to medical ICT with the increasing adoption of Electronic Health Records (EHRs); in the last decade, a lot of application have been developed in order to extract information and knowledge from electronic EHRs [1] [2]. In fact, when structured data is stored in an EHR, it is desirable to support automated systems at the point of care, and to help physicians in diagnosis. These studies endorsed most NLP applications in the medical field; for instance, those concerning the use of Twitter data and sentiment analysis to study diseases dynamics [3], or [4], where the correlation among "stress", "insomnia", and "headache" is analysed. In the field of medical application, the image processing are very useful in EHR data manipulation [5] [6], where medical images play an important role in particular to help physicians to monitor the evolution of complex pathologies [7]. In this

work a combination of image processing and NLP techniques are exploited to extract information from a scanned image of a medical prescription and analyze the semantics of the embedded information with the final goal of assessing its correctness according to the "medical request service" related to the prescription being examined. The system performs a classification in order to automatically authorize or not the medical service required within the prescription. Indeed, in Italy there exist a public medical assistance that provide free "Medical services". These though have to comply with certain parameters to be freely provided. Currently, the assessment of compliance with these parameters is manually performed by a proper operator. The proposed application aims to provide a mechanism to help the operator, or even replace his/her intervention. The proposed solution provides an user-friendly application to help the operator with a pre-analysis to isolate the few medical prescriptions that require a human operator to decide about their correctness, trying to automatize as many prescriptions as possible.

In Section II an overall description of the system with some implementation details is provided. Section III and IV describe respectively image pre-processing operations and text extraction. Section V discusses about the solution used in spelling correction and section VI describes the information classification task. Results are presented in section VII, while section VIII highlights conclusive remarks also outlining some future works.

## II. SYSTEM ARCHITECTURE

In this section we illustrate the proposed system and the solutions used to achieve the goal described in the introduction. As shown in fig. 1, the system is accessible via a Web application that works according to the following steps. First, the input image is examined to establish the type and format of the medical prescription that image represents, then, text is collected and corrected to further isolate and extract all relevant strings and the information based on previously collected strings is classified. Finally, the prescription is eventually considered as valid for further approval or not, according to specific criteria based on the information and related classification; we named these two possibilities as *grantable* and *not-grantable* respectively The ASP.NET framework has been adopted to develop the whole application; in particular,

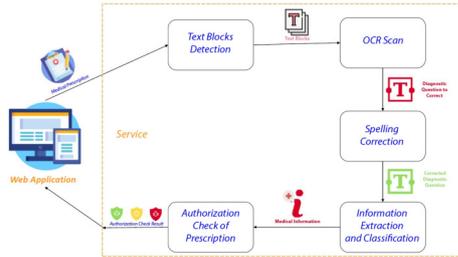


Fig. 1. Functional architecture of the proposed system

the ASP.NET Core version was considered, thanks to its support to open-source and multiplatform environments. In the paragraphs below, we discuss about each system module shown in fig. 1.

### III. MEDICAL PRESCRIPTION RECOGNITION

The first module of the proposed system aims to discriminate which type of medical prescription is being processed. The *type* is defined by the Italian national medical service ("Servizio Sanitario Nazionale", simply SSN in the following), indeed it includes:

- the prescription used to provide drugs, therapies, screenings or specialist examinations at the expense (entirely or partially) of the SSN; this prescription can be filled in by physicians that either works inside SSN structures (e.g., public hospitals) or they hold an agreement with SSN (being therefore a *partner* of SSN itself)
- the prescription where any medical care as those listed above are completely at the expense of the person that prescription was written for; in such cases, the physician is not required to hold any agreement with the SSN

The former type is also known as "ricetta rossa" (*red* prescription), and it is a specific prescription whose details also depend on local (region-based) rules, whereas the latter, known also as "ricetta bianca" (*white* prescription) has a general validity on the entire national territory, therefore in the rest of paper we just focus on this last type.

The system receives as input scanned images of prescriptions that must be classified as *white* ones or discarded. To accomplish this task, a supervised machine learning approach [8] is adopted. In particular, this well-known technique exploits a training set used to build a model that enables the classifier to perform the discrimination. Since we focus on white prescriptions only, the classifier is *binary*, i.e. it just establishes whether an image actually can be considered as a white prescription or not. To assess the effectiveness of the classifier, the widely adopted 75/25 approach has been considered as the training/validation set splitting. In addition, to prevent the overfitting problem, a data augmentation [9] has been performed on the dataset; note that due to its reduced dimension, we did not consider the cross-validation technique. The classifier we developed allows to detect white prescriptions with an effectiveness of about 93-95%. After that

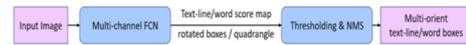


Fig. 2. EAST architecture

a white prescription has been acquired by the system, pre-processing steps [10] are carried out on the image in order to facilitate the subsequent OCR phase; in particular, we perform smart crop, gamma correction and image rotation.

The first operation is required since images present into our system are often simple photos provided by individuals that use their smartphone during the upload of the request of a medical service. Since the image is rarely provided by physician or other specialists, the accuracy of the prescription should be therefore improved in most cases by cropping the image to discard its negligible parts. Gamma correction [11] is usually performed in the process of digital imaging to restore as much as possible the lighting condition of the original image. Finally, image rotation is performed since it has been often generated via smartphone camera by standard users (SSN customers) so alignment could be required; to accomplish this, we exploit the barcode stored in the medical prescription, whose high contrast of black and white pixels allows a simple yet successful image alignment.

### IV. TEXT EXTRACTION

After image pre-processing, text is extracted to build a string dictionary where each entry represents a field of the medical prescription. This OCR phase is carried out using EAST [12] text detector in conjunction with Tesseract [13] free OCR software. EAST is the acronym of Efficient and Accurate Scene Text Detector, based on a multi-channel fully convolutional neural network (FCN) with an efficient pipeline, whose purpose is to isolate blocks of text embedded into an image. It can be schematized as in fig. 2, where the FCN produces both information about words/phrases recognition and about the geometry of the area that contains them; both are evaluated using a threshold-based mechanism that finally provide us with text blocks to be further processed via Tesseract. Using directly Tesseract to recognize the text contained in the scanned medical prescription provided unsatisfactory results for the data set used as input, reasonably due to the low quality of canned images provided by end users. For this reason, we first used EAST, whose effectiveness in isolating text blocks was higher, then passing each block to the OCR software; this approach revealed to be slightly lower but successfully for extracting data from prescriptions. In fig. 3 is represented a scanned white medical prescription (left side), together with main text fields extracted using EAST and Tesseract (right side).

The medical prescription contains some field that are relevant for our goal, in the following briefly described:

- 'regione' stands for region (administrative area Italy is splitted into), in this case with value 'Sicily'



Fig. 3. Text extraction from a sample medical prescription

- 'assistito' is the name of end user (registered to the SSN) the medical prescription refers to
- 'indirizzo', 'cap', 'citta' and 'provincia' are different parts of the end user's address
- 'cod\_fiscale' is the fiscal code (i.e. social security number) used to identify the user
- 'prescrizioni' is the list of medical services (e.g. drugs, therapies, screenings or specialist examinations as specified in section III); in the example shown, three blood tests are reported
- 'quesito' is the medical diagnosis as reported by the physician, that motivates the previous list of medical services

The last field 'quesito' is the most relevant for our purposes, since specific medical cares (field 'prescrizioni') can be allowed - and freely provided - by SSN only for specific diagnosis, therefore the field pair is used to rate the prescription as *grantable* or not, as discussed in previous sections.

## V. SPELLING CORRECTION

Once text has been extracted, we proceed with a spelling correction, that is required as in most OCR softwares residual errors still occur, in particular in our scenario where the quality of scanned images is not always high (as said previously) and also the text that appears in medical prescription is usually with a reduced font size. Furthermore, the recognition of the diagnosis block ('quesito' in fig. 3) is not trivial since this field is actually a free text with variable length, hence also spelling errors due to an incorrect entry by the physician are still possible. For all these reasons, the spelling correction is applied specifically to the diagnosis block; it consists of the following tasks:

- *Non-word* error detection, that is the detection of words characterized by incorrect spelling;
- *Isolated-word* error correction, i.e. the correction of the word written incorrectly without taking into account the surrounding context;
- *Context-dependent* error correction, that is word correction characterized by spelling mistakes based on the context.

Since no specific context is provided in the medical prescription, the spelling correction algorithm we implemented focus on the first and second task listed above. To fulfill its specification, the algorithm uses a words dictionary. This

vocabulary is preliminarily obtained by extracting all the words constituting the various rules used by the system in the classification phase.

ApplySpellingCorrection method, for each word in the diagnosis block searches in the vocabulary for all words that begin with the same letter. Then the TryCorrect method is called, passing as a parameter the set of words extracted from the vocabulary. This method leverages on the Damerau-Levenshtein distance [14] to accomplish its task; such a distance is the minimum edit distance between two strings, i.e. the lowest number of character insertion, removal and/or replacement to transform the former string into the latter.

If such distance is zero the word is correct, otherwise it must be replaced with the correct word in the dictionary. In our experiments, a threshold for such distance is chosen to limit the subset of candidate words extracted from the dictionary. In the case of a null subset (for the given threshold), the word to replace is considered *unknown*, since no proper word in the dictionary has been found. The higher the threshold, the more (and possibly not suitable) words will form the subset of candidates, hence keeping as lowest as possible the value is recommended; we carried out successfully experiments using '1' as threshold.

The performance obtained from the Spelling Correction algorithms are quite satisfactory, this is due not only to the efficiency of the algorithm, but also to the use of the cache, in which all the rules are stored, the word vocabulary used for spelling correction and the different weights used to calculate the score of the different rules.

## VI. INFORMATION CLASSIFICATION

The goal of information classification is to assess whether a given prescription is grantable or not, as specified in previous sections; to do this, the text extracted (and eventually corrected) is properly classified exploiting both the *Syntactic rules* and the *Rule-based tagging* NLP technique [15]. Syntactic rules are used to model all valid grammar sequences, whereas Rule-based approaches use contextual information to assign tags to unknown or ambiguous words (often called *context frame rules*). Rule-based taggers generally require supervised training, but also other approaches are available [16].

The proposed solution exhibits simplicity and good performance as it only requires the use of syntactic rules for pattern matching information extraction, and the use of rules that use data belonging to the context frame, to extract new categories of information. In this first stage of development rules and patterns have been manually built, but this time-consuming and error-prone task is going to be removed by automatized rules generation in further development of this work. The well know schema for a syntactic rule contains three tags, i.e. Source, Target and Data (see eq. 1).

$$Source \Rightarrow Target \# Data \quad (1)$$

The *Source* attribute indicates a specific pattern that the system must detect within the text string being analyzed before

it can apply the rule itself. The pattern for this attribute can be either a simple string or a syntactic expression to specify that a regular expression must be matched to detect the pattern within the text. To discriminate the type of pattern the *Source* attribute is set, rules are classified in *Regex Rule* and *String Rule*.

In a *Regex Rule* the *Source* attribute contains a Placeholder whose structure is shown in 2 and 3, where *placeholder\_value\_1* and *placeholder\_value\_n* are correct pattern matching expressions.

$$\{\{placeholder\_type : *\}\} \quad (2)$$

$$\{\{placeholder\_type : placeholder\_value_1|placeholder\_value_n\}\} \quad (3)$$

The *Target* attribute indicates the Placeholder that must be used when applying the rule to replace the pattern indicated by the *Source* attribute detected in the analyzed text. This attribute can be set in two different ways. The former requires that it simply contains the string to be used to perform the replacement, whereas the latter requires that it contains the index indicating the position of the word contained within the *Source* attribute to be used as a Placeholder when the rule is applied. In order to distinguish the two set modes, the index is always preceded by the special character £(this solution was chosen to avoid redundancy in rule coding). The replacement of pattern matching present in the rule with the value contained in the *Target* field is a text tagging operation hence the related placeholder must be characterized by a tag structure. In particular, the system provides that the placeholder of each rule is enhanced by a string having the following structure:

*placeholder\_type : placeholder\_value*

Finally, the *Data* attribute in eq. 1 indicates all the information that can be extracted from the analyzed text when the rule is applied. In general, this attribute is enhanced by a string consisting of the representation of information in JSON format. The system also allows this string to contain parameters represented in two different ways, based on their semantics. In a first case the parameter is indicated by an integer preceded by the special character &, where the integer indicates the position within the *Source* attribute of the word to be used to evaluate the parameter; another option is to use the following syntax:

*index.attribute\_name*

where *index* indicates the position within the *Source* attribute i.e. the key that must be used to access the data structure maintained by the system containing all the information extracted through the application of the rules and *attribute\_name* instead indicates the category of information of interest.

The syntax used for the coding of rules does not require that the *Data* attribute must necessarily contain a string consisting of the representation of information in JSON format. In fact, it is possible to associate the *Data* attribute with a string having the following structure:  $index_0 + index_1 + \dots + index_n$

When the *Data* field receives such a value, information extracted when the rule is selected are collected from those having indexing keys equal to words from the *Source* field at position  $index_0, index_1, \dots, index_n$ .

The algorithm to classify information is implemented in C# and operates as follows. After rules are fetched from the configuration file, it rates each rule with a score depending on the number of words in the *Source* field, distinguishing placeholders from simple strings. In particular, the *placeholder\_type* of each placeholder in the *Source* lead to a different weight for its related placeholder; such weights can be manually specified within the configuration file. Rules rating allows to establish their application order (priority).

## VII. SYSTEM TESTING

In this section we describe the testing phase carried out to evaluate the performance of spelling correction and information classification algorithms, with the final purpose of establishing whether an input medical prescription can be classified as grantable or not, or eventually whether the system was not able to classify it at all.

To this purpose, a dataset with about 800.000 text rows coming from medical prescription has been used, while the rule set for the classification contains about 5000 mapping rules; note that in this test we did not consider strings extraction from medical prescription, since we focused on the assessment of performance classification. In order to efficiently perform the test, a C# script working in parallel for each row invoking an HTTP POST at the REST service was developed. During the test, several information are collected: number of traumas, number of diagnostic query, symptoms and areas present in it and, moreover, the number of spelling corrections (incorrect words and their correction and the Damerau-Levenshtein distance).

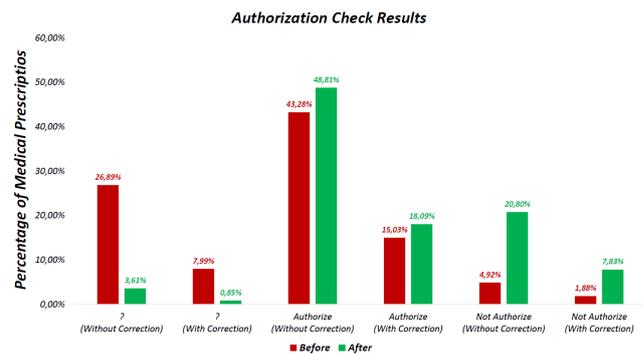


Fig. 4. Test Result

In figure 4 are shown the test results, where 'Authorize' and 'Not Authorize' indicate the grantable property, and '?' label collects those prescriptions that the system was not able to classify (named *unclassifiable* in previous sections. For each case, the two assessment (when spelling corrections are applied or not) are indicated separately.

The red line shows that approximately 30% of prescriptions are unclassifiable, a relevant (and then unacceptable) percentage in particular for the case when spelling corrections were not required. To tackle this situation, a first step was to map a lot of terms by writing a rule for each one of them, but this operation did not lead to a significant performance improvement, so we decide to include two additional information categories:

- *\_si*: featuring all those words whose combination, if present within the diagnosis text block, make the prescription grantable.
- *\_no*: where all those words that do not affect the classification of the medical prescription at all are stored.

The green line in figure 4 shows the result of the classification for this improved solution; in this case the classification improves significantly since only the 5% of medical prescriptions are considered unclassifiable.

### VIII. CONCLUSIONS

The main goal of the proposed system is to develop a service for the analysis and authorization of medical prescriptions. Results shown that in most cases the system allows automatic classification (as grantable or not) and only 5% were not automatically classified; from tests carried out on 800,000 recipes only around 4000 therefore required manual operator intervention. The classification phase is the most relevant part of the proposed system and its quality strictly depends on the number of rules used. Their writing is a time-consuming and error-prone task, especially if manually built, therefore a planned further work is to exploit machine learning techniques to automatically manage the set of rules.

### ACKNOWLEDGMENT

This work has been developed in cooperation with Pre-  
vimedical s.p.a. and Previnet s.p.a

### REFERENCES

- [1] V. Carchiolo, A. Longheu, M. Malgeri, and G. Mangioni, "Multisource agent-based healthcare data gathering," in *Proc. of FedCSIS*, Sep. 2015, pp. 1723–1729. [Online]. Available: <https://doi.org/10.15439/2015F302>
- [2] Y. Si and K. Roberts, "A frame-based nlp system for cancer-related information extraction." *AMIA Annu Symp Proc*, vol. 2018, pp. 1524–1533, 2018.
- [3] V. Carchiolo, A. Longheu, and M. Malgeri, "Using twitter data and sentiment analysis to study diseases dynamics," in *Proceedings of ITBAM 2015*, vol. 9267. New York, NY, USA: Springer-Verlag New York, Inc., 2015, pp. 16–24. [Online]. Available: [http://dx.doi.org/10.1007/978-3-319-22741-2\\_2](http://dx.doi.org/10.1007/978-3-319-22741-2_2)
- [4] S. Doan, E. W. Yang, S. Tilak, and M. Torii, "Using natural language processing to extract health-related causality from twitter messages," in *IEEE ICHI-W*, June 2018, pp. 84–85. [Online]. Available: <https://doi.org/10.1109/ICHI-W.2018.00031>
- [5] S. A. Parah, J. A. Sheikh, F. Ahad, N. A. Loan, and G. M. Bhat, "Information hiding in medical images: a robust medical image watermarking system for e-healthcare," *Multimedia Tools and Applications*, vol. 76, no. 8, pp. 10 599–10 633, Apr 2017. [Online]. Available: <https://doi.org/10.1007/s11042-015-3127-y>
- [6] B. Shickel, P. J. Tighe, A. Bihorac, and P. Rashidi, "Deep ehr: A survey of recent advances in deep learning techniques for electronic health record (ehr) analysis," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 5, pp. 1589–1604, Sep. 2018. [Online]. Available: <http://dx.doi.org/10.1109/JBHI.2017.2767063>
- [7] G. Litjens, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60 – 88, 2017. [Online]. Available: <https://doi.org/10.1016/j.media.2017.07.005>
- [8] S. B. Kotsiantis, "Supervised machine learning: A review of classification techniques," in *Proc. of the 2007 Conf. on EAIACE: Real World AI Systems with Applications*. Amsterdam, The Netherlands, The Netherlands: IOS Press, 2007, pp. 3–24. [Online]. Available: <https://dx.doi.org/10.1007/s10462-007-9052-3>
- [9] S. C. Wong, A. Gatt, V. Stamatescu, and M. D. McDonnell, "Understanding data augmentation for classification: When to warp?" in *2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, Nov 2016, pp. 1–6. [Online]. Available: <http://dx.doi.org/10.1109/DICTA.2016.7797091>
- [10] W. Bieniecki, "Image preprocessing for improving ocr accuracy," in *Intl. MEMSTECH Conf.*, 06 2007, pp. 75 – 80. [Online]. Available: <https://dx.doi.org/10.1109/MEMSTECH.2007.4283429>
- [11] X. Guan, "An image enhancement method based on gamma correction," in *2nd Intl. Symp. on CID*, vol. 1, Dec 2009, pp. 60–63. [Online]. Available: <https://dx.doi.org/10.1109/ISCID.2009.22>
- [12] X. Zhou, C. Yao, H. Wen, Y. Wang, S. Zhou, W. He, and J. Liang, "EAST: an efficient and accurate scene text detector," *CoRR*, vol. abs/1704.03155, 2017. [Online]. Available: <https://doi.org/10.1109/CVPR.2017.283>
- [13] Tesseract, "Tesseract Open Source OCR Engine," <https://github.com/tesseract-ocr/tesseract>, last accessed 08 May 2019.
- [14] F. J. Damerau, "A technique for computer detection and correction of spelling errors," *Commun. ACM*, vol. 7, no. 3, pp. 171–176, Mar. 1964. [Online]. Available: <http://doi.acm.org/10.1145/363958.363994>
- [15] E. Brill, "A simple rule-based part of speech tagger," in *Proceedings of the Third Conference on Applied Natural Language Processing*, ser. ANLC '92. Stroudsburg, PA, USA: Association for Computational Linguistics, 1992, pp. 152–155. [Online]. Available: <https://doi.org/10.3115/974499.974526>
- [16] E. Brill and M. Pop, *Unsupervised Learning of Disambiguation Rules for Part-of-Speech Tagging*. Dordrecht: Springer Netherlands, 1999, pp. 27–42. [Online]. Available: [https://doi.org/10.1007/978-94-017-2390-9\\_3](https://doi.org/10.1007/978-94-017-2390-9_3)



# 1<sup>st</sup> Workshop on Data Analysis and Computation for Digital Ecosystems

**T**HE progress in Information and Communication Technologies (ICT) builds a strong basis to realize cross-domain data computation approaches. However, the data flow, coming from local sensors up to satellite remote sensing, business processes, enquiries, social networks and other manifold sources is extremely complex and dynamic. Therefore, sufficient data conditioning as well as analysis methods and corresponding ICT infrastructures are crucial for the extraction and handling of the information embedded in the data. Especially the development and implementation of efficient workflows for the integration of heterogeneous data types across various stakeholders presents a big challenge in the development of data-driven smart services. New paradigms are needed for empowering science, society and business to understand and design the interaction of systems and processes on different spatial and temporal scales. The idea of the workshop enables an interdisciplinary discussion of challenges and potential solutions in the new world of cross-domain data integration and information extraction (e.g. energy, health, environment, economy). We invite contributions presenting challenges, solutions, cases studies, and best practices dealing with ecosystem management, big data and workflow modeling. Competencies from the field of information systems, data science, environmental science, computer science, mathematics, medicine, energy and business are required to develop new promising ideas that contribute to the investigation and extraction of information from cross-domain data.

This special session is a joint event of WIG2, the Scientific Institute for Health Economics and Health Services Research, the Institute for Information Systems of the University of Leipzig and the Helmholtz Center for Environmental Research—UFZ.

## TOPICS

The topics of interest therefore include but are not limited to:

- Ecosystem design and management for cross-domain data exploitation
- Cross-domain maturity in businesses of different sizes and industry sector

- Case studies investigating challenges and potentials of cross domain analytical
- Efficient data management approaches along the chain of information
- Fusion of data, methods and results
- Data quality assurance
- IT-infrastructures and -architectures
- From data to variable and from variable to indicator workflows for smart services
- Best practices to improve interoperability
- Dealing with privacy and data ownership

## EVENT CHAIRS

- **Bumberger, Jan**, Helmholtz-Centre for Environmental Research – UFZ, Germany
- **Alt, Rainer**, University of Leipzig / Social CRM Research Center, Germany
- **Dietrich, Peter**, Helmholtz-Centre for Environmental Research – UFZ, Germany
- **Franczyk, Bogdan**, University of Leipzig, Germany
- **Reinhold, Olaf**, University of Leipzig / Social CRM Research Center, Germany

## PROGRAM COMMITTEE

- **Bosch, Jan**, Chalmers University of Technology, Sweden
- **Cirqueira, Douglas**, Dublin City University
- **Fileto, Renato**, Federal University of Santa Catarina, Brazil
- **Kabisch, Nadja**, Humboldt-Universität zu Berlin, Germany
- **Kirn, Stefan**, University of Hohenheim, Germany
- **Kirsten, Toralf**, Mittweida University of Applied Sciences, Germany
- **Lobato, Fábio**, Federal University of Western Pará, Brazil
- **Rot, Artur**, Wroclaw University of Economics, Poland
- **Suomi, Reima**, University of Turku, Finland
- **Veijalainen, Jari**
- **Viana, Julio**, Social CRM Research Center



## Location Intelligence in Cogenerated Heating Potential Data Analysis

Almir Karabegovic  
University of Sarajevo,  
Faculty of Electrical  
Engineering, Kampus  
Univerziteta u Sarajevu,  
71000 Sarajevo, BiH  
Email:  
akarabegovic@etf.unsa.ba

Mirza Ponjavic  
International Burch  
University, Sarajevo,  
Francuske revolucije bb,  
Ilidža 71210, Bosnia and  
Herzegovina  
Email:  
mirza.ponjavic@gis.ba

Neven Duic  
University of Zagreb,  
Faculty of Mechanical  
Engineering and Naval  
Architecture, Ivana Lučića 5,  
10002 Zagreb, Croatia  
Email:  
neven.duic@fsb.hr

Tomislav Novosel  
North-West Croatia  
Regional Energy  
Agency, Andrije Zaje  
10, 10000 Zagreb,  
Croatia  
Email:  
tnovosel@regea.org

**Abstract**—Different methodologies are used to assess the potential for using high efficiency cogeneration for cooling and heating. They are mostly adapted to the availability of data and tools for their analytical processing. This paper presents the approach applying location intelligence as a tool that allows using geospatial analysis algorithms and geovisualization of its results. Due to the extremely large amount of data and the dependence of the results on their accuracy and the level of aggregation, the initial methodology of the analytical process implied two steps: wide scale mapping by the "top down" method, and local mapping by "bottom up" method. However, in order to overcome the problem of regional disparities of quality and the existence of spatial data, certain adaptations of the initial methodology have been made considering the need for a single analytical approach for the entire area of interest. Randomized control of the obtained results indicate that applied geospatial algorithms satisfy the required level of accuracy and reliability of the final methodology.

### I. INTRODUCTION

IN 2004, the European Parliament and the Council of the European Union adopted the Directive 2004/8 EC whose purpose is to increase energy efficiency and develop high efficiency cogeneration of heat and power. The Annex III of the Directive defines the "High Efficiency Cogeneration" as the cogeneration production from cogeneration units that provides primary energy savings (PES) at least 10% compared with the references for separate production of heat and electricity (small scale cogeneration units, installed capacity below 1 MWe, and micro cogeneration units, installed capacity below 50 kWe, must provide primary energy savings, that is  $PES > 0$ ). Bosnia and Herzegovina (BiH) signed the Treaty establishing the Energy Community calling for the adoption and implementation of Acquis Communautaire on energy, environment, competition and renewables. The Acquis' essential directives pre-stipulate the area of energy end-use efficiency and energy services, energy performance of buildings and labelling. Bosnia and Herzegovina also signed to fulfill the obligations under the latest EED directive 2012/27EU [1] including the need to adopt policies incorporating local and regional potentials for using efficient heating and cooling systems, in particular

those using high-efficiency cogeneration and the potential for developing local and regional heat markets. According to the directive each country shall carry out a comprehensive assessment of the potential for the application of high-efficiency cogeneration and efficient district heating and cooling. The assessment covers entire territory of the country taking into account its specificities like present situation with central heating and cogeneration systems, climate conditions, economic feasibility and technical suitability on potentials for the application of high-efficiency cogeneration and efficient district heating and cooling. The analysis used in the assessment facilitate the identification of the most resource-and cost-efficient solutions to meeting heating and cooling needs [2]. This type of analysis has already been to some extent implemented within the European Union through, for example, the Multi-level Actions for enhanced Heating & Cooling Plans (STRATEGO) projects funded through the Intelligent Energy Europe [3] and Heat Roadmap Europe (HRE) [4].

Within the technical assistance provided by GIZ BiH (Deutsche Gesellschaft für Internationale Zusammenarbeit), it was carried out the study by a consultant group which, as part of the assessment, resulted in a mapping approach using geoinformation system (GIS) tools for presentation demands and potentials of using thermal energy.

This paper describes the methodology and the approach applying location intelligence (LI) as a tool that allows using geospatial analysis algorithms and geovisualization of consumption of thermal energy and the potential for the application of high-efficiency cogeneration and efficient district heating and cooling.

### II. THE METHODOLOGY AND THE MAPPING APPROACH

In order to properly assess the potential for the utilization of highly efficient heating and cooling technologies, primarily district systems and renewables, the spatial distribution of the demand and potential supply must be analyzed. Such an assessment demands a great deal of data which is often not available or at least not public or in a usable format. To work around most of these issues a two-

step approach is suggested here. In the initial step, the annual heat demand gathered on a national, regional or municipal level will be distributed spatially according to parameters such as population or building density. This will be used to perform a rough estimate to determine which areas are deserving of a more detailed analysis based on heat demand density and/or availability of local renewable resources. The second step will focus on the areas which have demonstrated a high potential for the utilization of highly efficient heating and cooling technologies either because of a high demand density or due to the abundance of renewable energy sources [5].

#### *A. Step one – top down analysis*

As it has been mentioned above, the initial mapping step will be conducted with a top down approach. Annual heating and cooling demand data will be gathered on national, regional and municipal levels (the availability and quality of data will determine the aggregation level). The collected data will be spatially distributed according to parameters such as population densities and land coverage. These data can be found in georeferenced forms in population and building censuses, cadasters and public databases such as the CORINE land use map [6]. The resolutions of these data are varied from very precise such as individual buildings or raster of 100 by 100 meters to less precise ones such as municipal or county level. The end result of this step will be a heating and cooling demand GIS map with a resolution of at least 1 by 1 km. Additionally to the demand, potential supply sources will also be analysed here. Potential heat sources such as excess industry heat, solar and geothermal energy as well as waste will be evaluated and presented in a GIS form. This analysis will provide the possibility to perform an initial evaluation of areas suitable for the exploitation of highly efficient heating and cooling technologies. Areas deemed worthy of further, more detailed, evaluation will be processed in the second step. This will include mostly larger cities and areas close to industrial parks and sources of renewable heat [5].

#### *B. Step two – bottom up analysis*

Areas identified as worthy of further evaluation will be analysed in more details in the second step. Here, a bottom up analysis will be implemented where the scale of the analysis will be individual buildings. For this purpose, cadaster data will be utilized to identify the locations of buildings and other vital data, where available, such as area, volume, age, use, level and time of refurbishment and so on. Representative buildings will be used or modelled to calculate the specific heating and cooling demands for certain building types based on parameters such as use, age and level of refurbishment. This will result in demands per area which will then be used to calculate the heating and cooling demand of all observed buildings. Where digital cadaster data is not available the CORINE land use map will be used and the process will result in a 100 by 100 meter

map of the area. Additionally, existing district heating infrastructure will be mapped. This will include production facilities and distribution grids. The resulting map will include the heating and cooling demands, existing infrastructure and potential sources of local renewable and excess heat [5].

Heating and cooling demands represent value (heat) energy required by population for heating and cooling buildings and spaces inhabited.

#### *C. The mapping approach*

The analysis of this type requires an extremely large amount of data, and the results depend on their accuracy and aggregation level. The methodology described above consists of two steps: state mapping by the "top down" method, and local mapping by "bottom up" method.

The first step of this process allows the development of a coherent map of demand for heating and cooling and potential sources of waste heat for the entire observed area of Bosnia and Herzegovina, its entities of the Federation of Bosnia and Herzegovina (FBiH), Republika Srpska (RS), and Brcko District (DB). The results obtained should serve to determine the priority areas for further analysis in the second step, which is based on a detailed analysis of demand at the level of individual facilities. This would provide detailed maps for areas of potential interest for the exploitation of district heating and cooling systems and high efficiency cogeneration which represented an important source of data for all further analyzes.

This approach would first result in a state map in resolution of 1x1 km, and as a final result, it would be obtained a set of maps with priority areas in 100x100 meter resolution.

Generally, this concept implies the approach applying location intelligence as a tool that allows using geospatial analysis algorithms and geovisualization of its results. In order to realize it may require the application of various tools, spatial data integration procedures [7], types of analysis such as geometric, topological, set oriented, grid and graph analytic methods [8][9].

### III. GEOSPATIAL ANALYSIS OF COGENERATED HEATING POTENTIAL DATA

In many instances the process of spatial analysis follows a number of well-defined (often iterative) stages: problem formulation; planning; data gathering; exploratory analysis; hypothesis formulation; modeling and testing; consultation and review; and ultimately final reporting and/or implementation of the findings. GIS and related software tools that perform analytical functions only address the middle sections of this process [10].

In this study, the first two stages related to problem formulation and planning are defined by a project assignment, so that the issues of the available data, the

TABLE I.

LISTS OF DATA SETS USED FOR TOP DOWN AND BOTTOM UP ANALYSIS

Data set	Resolution	Time reference	Source
Census data / population	Enumeration area	2013	Institutes for statistics in BiH
Census data / housing	Enumeration area	2013	Institutes for statistics in BiH
Climate zone / polygons	Country	2018	Meteorological institutes in BiH
Enumeration areas / polygons	Approx. 100 by 100 m	2013	Institutes for statistics in BiH
Municipal boundaries	Cadaster municipality	2018	Geodetic administration of BiH entities
Cadastral data / buildings	Cadaster parcel	2018	Geodetic administration of BiH entities
CORINE Land Cover	25 ha	2012	European Environment Agency
OpenStreet Map	-	2018	OS data
Sources of waste heat / points	-	2018	Institutes for spatial planning in BiH
Industrial facilities / points	-	2018	Institutes for spatial planning in BiH
Greenhouse gas emissions	-	2016	Third national communication report on greenhouse gas emissions of BiH
Orthophoto maps	Scale 1:2500 / 1:5000	2012	Geodetic administration of BiH entities
District Heating Systems Infrastructure	Cadaster of infrastructure facilities	2018	District heating companies in BiH
Meteorological data	Municipality	2018	Meteorological institutes in BiH

analytical process itself, and the testing of the applied model, ie control of the results are outlined below. The quality and completeness of the data can have a dramatic impact and can cause implications for the results and the process of spatial data analysis itself [11]. During the geospatial analysis it was concluded that the sets of available data were not enough to carry out the analysis according to the initial methodology, and its adaptation was proposed. Also, a brief overview of the results of the analysis has been described, showing how well adapted the methodology is acceptable for further application.

### A. Available data

The major challenge for the design of the heating potential assessment study was the availability and quality of input data in various formats and for various purposes. For the purpose of providing data for calculating the needs for heating and cooling, various data sources have been used: land cadaster, population census, weather statistics, geodetic surveys, administrative statistical spatial boundaries, address register, map data (climate, rainfall, soil, forest cover), land use and land cover data CORINE 2012, official orthophotos, DEM, LiDAR, and public information on the Internet, such as studies on construction types, remote heating, energy efficiency, climate change and other sources related to the BH area. Useful sources were also typologies of residential and public buildings in BiH [12][13], as well as cogenerate sources and waste material co-incineration data [14].

Among the most important available data sets are the census data related to housing units, the way of heating and energy consumption. Some of the attributes associated with these sets were: the area of the apartment, the type of heating, the type of building, the year of construction, the number of floors, the building material, the building structure and the state of the object. Also, data for all existing district heating systems in BiH are collected.

Further, various data was collected for the needs of determining the specific annual heat energy required for heating and the annual cooling energy required per sector, which includes: altitude for municipalities, external temperature for building design, degree of development of municipalities, number of inhabitants per municipality, number of employees per sector, number households by municipalities, average household size per municipality, and total average residential area (usable area) per capita. Table I lists some of the spatial data sets used for the analysis.

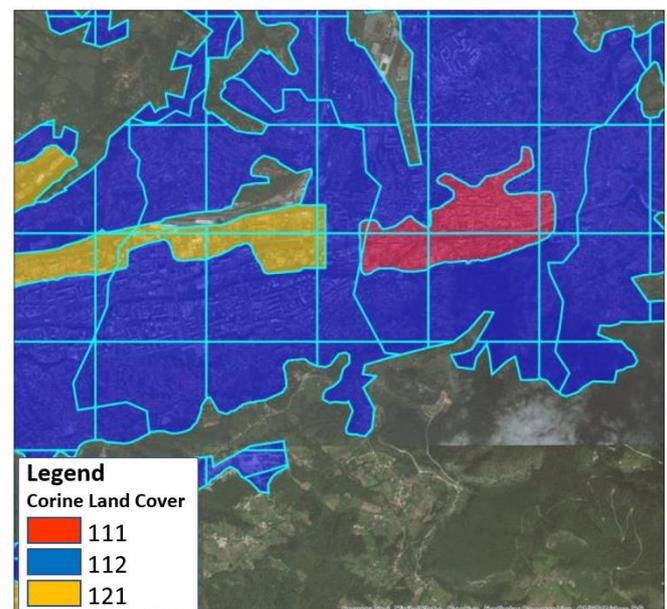


Figure 1. Intersection network, municipality boundaries and CLC data sets

### B. Geospatial analysis of heating potential data and adaptation of the methodology

In order to fully apply the described methodology, in addition the aforementioned sets, many more data were needed, but some were not available in useful form. Particular problems were the lack of geometric data on residential and industrial buildings in some parts of the country, as well as topological defects related to polygons of statistical units, which are used for the intersection with the network of squares for the presentation of the heating demands. Also, the lack of comprehensive and consistent data for the non-residential sector in BiH was particularly evident.

In order to overcome all the problems of regional disparities of quality and the existence of spatial data, certain adaptations of the initial methodology have been made considering the need for a single analytical approach for the entire area of interest.

#### 1) Country level mapping by the "top down" method

One of the aspects of this analysis was the determination of heating and cooling needs at the national and local level. In the first phase (top down) is made calculation of energy needs on state level for the following sectors: housing, public and service sector, and industry (Figure 2).

The procedure is based on determining the total area of housing, the public and service sectors, and the industry sector (in km<sup>2</sup>). This activity was carried out using the CORINE database, that is CORINE Land Cover (CLC), where are filtered only the categories of land cover which exclusively belonging to the one of the following classes:

- Continuous urban fabric (CLC code 1.1.1),
- Discontinuous urban fabric (CLC code 1.1.2.),
- Industrial or commercial units (CLC code 1.2.1).

The processing continued to provide information on how

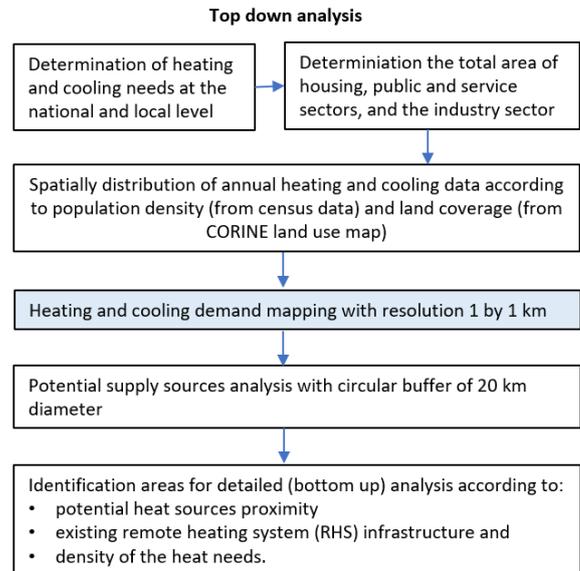


Figure 2. Steps of top down analysis

many municipalities have certain categories of land (in km<sup>2</sup>). Further, the same information is displayed spatially by a 1x1 km resolution raster. By addressing these data with data on energy needs calculation, final data was obtained at the municipal level and at the level of each square (1x1 km).

For this purpose, it was necessary to perform several operations in the GIS to enable the correct spatial distribution of the energy demand data. The first step involved intersecting three sets of data:

- Corina Land Cover (CLC) with
- 1x1 km network; and
- the boundaries of the municipalities.

Figure 1 gives the example of the data intersection. In this way, they are determined the area of each particle obtained, its belonging to the certain municipality and square of the

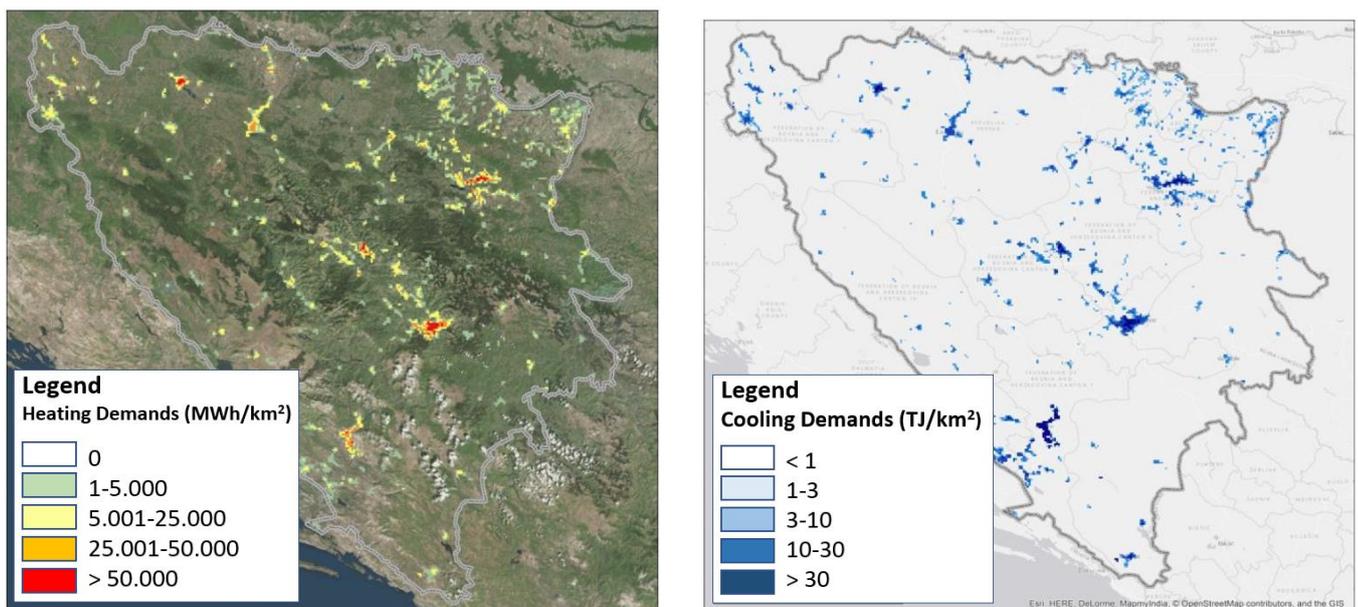


Figure 3. Maps of heating (left side) and cooling demands (right side) in Bosnia and Herzegovina with values in MWh/km<sup>2</sup> and TJ / km<sup>2</sup> units per year

network, and then heating and cooling demand are allocated to each particle.

Also, percentage share (%) of each particle in total area of a given municipality is determined for each purpose. The required annual heating energy for residential, public and service sector ( $RED_{pR/PS}$ ) is calculated using the formula (1):

$$RED_{pR/PS} = \frac{A_{pi}}{A_{in}} \times RED_{nR/PS} \tag{1}$$

where are:

- $A_{pi}$  area of the particle „p“ with purpose „i“;
- $A_{in}$  total area of land belonging to CLC class for purpose „i“ in certain municipality „n“;
- $RED_{nR/PS}$  required heating energy in municipality „n“ for residential, public and service sector.

In this way, the required heating energy for the industry sector is also determined ( $RED_{pl}$ ).

The total required heating energy at the level of each square should correspond to the sum of the individual needs of each respective particle for the different purposes of the CLC class land.

The same principle was used to determine the required cooling energy.

The applied methodology resulted in distribution of heating and cooling requirements at the level of the network square (1 km<sup>2</sup>) for the country. Due to a clearer view, two types of maps were created (Figure 2). with different measuring units (in MWh / km<sup>2</sup> and in TJ / km<sup>2</sup> yearly).

Figure 3 Maps of heating (left side) and cooling demands (right side) in Bosnia and Herzegovina with values in MWh/km<sup>2</sup> and TJ / km<sup>2</sup> units per year

For the purpose of allocating potential sources of waste heat, 21 sources, which are mainly large industrial and energy facilities (Figure 4), are identified and mapped. Currently, much of the surplus energy from these plants,

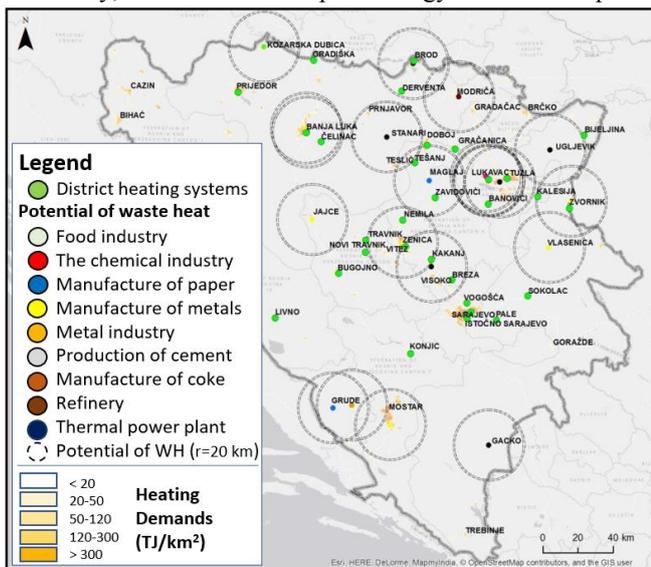


Figure 4. An overview of heating needs (TJ / km<sup>2</sup>) and waste heat potential

which can be used for heating and cooling purposes, remains unused, although these plants are mostly situated near densely populated settlements. The heat surplus from these plants can be theoretically estimated based on primary energy consumption or CO<sub>2</sub> emissions. A detailed description of the methodology for calculating the waste heat from CO<sub>2</sub> emissions can be found in the STRATEGO [15] report and research works [16].

For the purposes of assessing the potentials of high efficiency cogeneration in BiH, it is made the selection of areas which have relatively high density of heating needs, and these are mostly urban parts of municipalities and cities.

2) Local level mapping by the "bottom up" method

The second step relates to the mapping of heating and cooling demands at the local level for every selected area (municipality or city). Similarly, to the first step of the mapping (top down), the calculation for energy needs at the level of individual municipalities has been implemented separately for housing, public and services sector and for industry sector (Figure 5).

For selection of municipalities and cities, the following criteria were applied:

- existing remote heating system (RHS) infrastructure,
- proximity of the waste heat potential and
- the density of the heat needs.

The illustration of these data is shown in Figure 4.

According to the first and key criterion it is needed to select the urban environment with existing remote heating. The second criterion is the potential for exploitation of waste heat from industrial plants, as waste heat is the most convenient source for remote heating. A radius of 20 km has been set for areas with potential for waste heat utilization. The third criterion is the density of energy needs that goes to

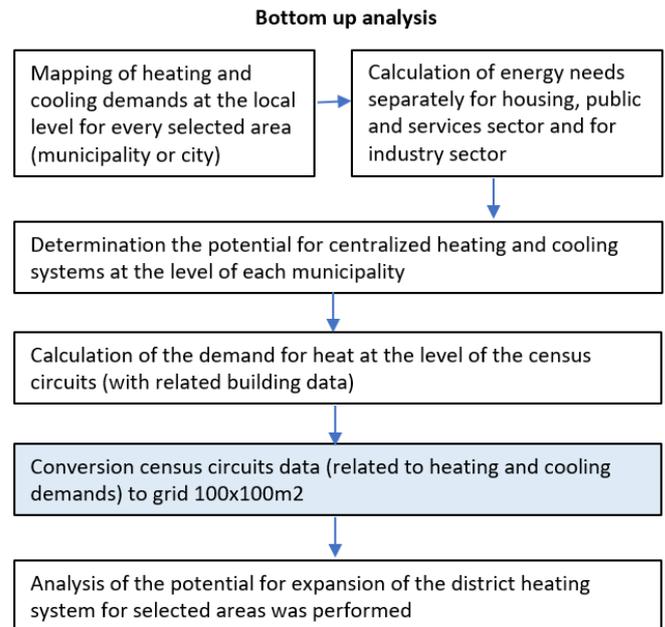


Figure 5. Steps of bottom up analysis

favor of those urban environments that do not have existing district heating and are not close to an industrial plant with waste heat but can justify investment because of high demands.

Taking into account the aforementioned criteria, a total of 39 cities and municipalities have been selected for further analysis, representing over 60% of total heating needs and 55% of cooling demand in Bosnia and Herzegovina.

According to the adopted methodology, heat energy needs for selected areas are determined for the residential and non-residential sector separately, and the results are presented on maps by network in resolution 100 x 100 m (Figure 6a).

In the final part of the study, it was determined the potential for centralized heating and cooling systems at the level of each municipality. Also, the analysis of the potential for expansion of the district heating system for selected areas was performed (Figure 6b). In discussion section of this paper, the deficiencies related to the application of the methodology are explained.

### C. Control of results

By visual detection, it is obvious that heating and cooling demands are distributed in terms of expectations and according to the areas where the largest number of inhabitants is concentrated (population per km<sup>2</sup>) and depending of geographic and climatic characteristics. In order to validate the data, it is conducted a detailed analysis of the municipalities of the City of Sarajevo (Figure 7) as well as several other municipalities, using information and experience gained through work on previous similar projects.

Based on the data presented in Figure 7, it can be concluded that the data spatial distribution satisfies, and the data are geocoded in accordance with the expectations and the real situation. The figures in the quadrants indicate heating demands in MWh / km<sup>2</sup>. The areas of the Novi Grad municipality (Figure 6a), namely the settlements Alipasino Polje, Hrasno and Cengic Vila indicate the high demand in this case as in the reality. On the other hand, the values

related to the peripheral parts of the city in Figure 6b, due to the smaller population density and the construction of individual residential buildings, indicate less heating needs. It can be concluded that the data is logically distributed in the space. For the purpose of final validation, a detailed analysis of the values based on their random selection was performed in accordance with the statistics norms.

For validation purposes, these results are also compared to other available data. For example, in the UK value of the largest demand quadrant is over 200,000 MWh, while in BiH it is close to 90.00 MWh. Considering the urbanization and population density (e.g. population density in the urban area of London is 1.65 times higher than in the urban area of Sarajevo), the data obtained can be considered as credible. Analyzing the available data from other countries in Europe (Spain, Czech Republic, Poland, Luxembourg), the results satisfy the expected framework (by the benchmark analysis).

## IV. DISCUSSION

Location intelligence (LI) implies an insight into the geospatial relationships between the phenomena that are studied to solve the problem important for spatial decision making [17]–[21]. This allows a layered spatial display of data sets processed using GIS tools for their transformation, analysis, and visualization.

In this study, LI had a twofold role:

- to prepare the cartographic basis for complex spatial analysis
- to visualize the spatial distribution of demands and potentials for the use of heat energy from cogeneration.

During the processing of geometric and attribute data in GIS special challenge was the heterogeneity of data (both syntax and semantic) [22] [23]. Namely, part of the accessible data represented attribute dwelling data referenced by census numbers and geometric data (polygons) of census circles that came from different sources. Due to its consistency and the unique model, it has been suggested to

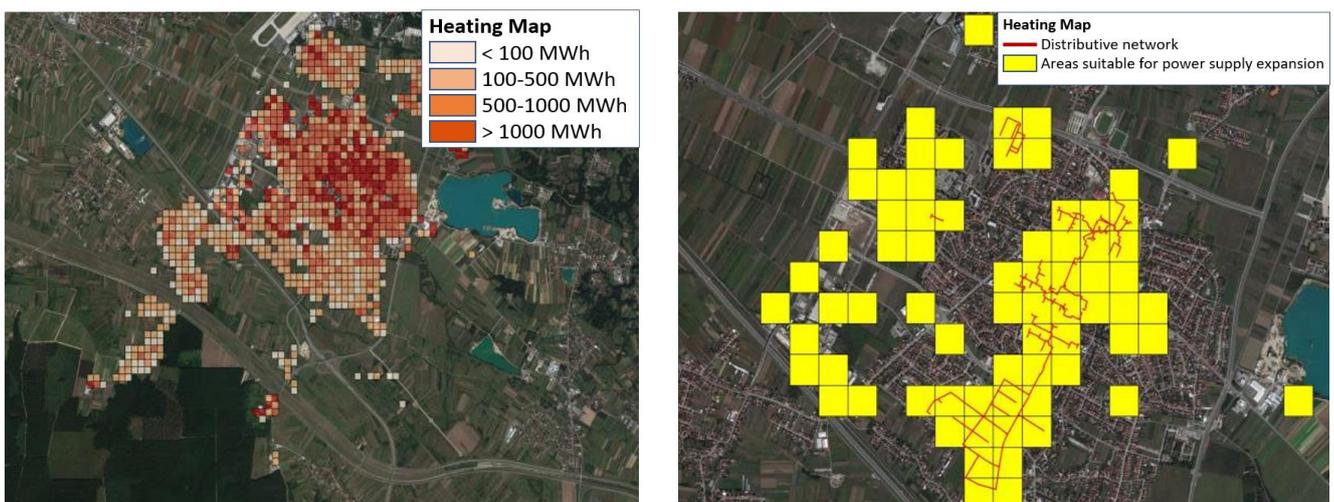


Figure 6. Examples of maps: a) heating demands presented by network 100 x 100 m (left side) and b) potential for expansion of remote heating system (right side)



Figure 7. Heating demands in MWh/km2 at 100 x 100 m network level in the City of Sarajevo, municipalities of Centar and Stari Grad

apply these data for the whole BiH area. Thus, they were used as a basis for the overall analysis, including the calculation of the demand for heat at the level of the census circuits.

Certain geometric data sets (for example, Brcko District) did not meet the topological conditions, so their further processing was done, including topology control, polygonization, and redrawing.

Further, the data on spatial content objects intended for commercial and other non-residential needs were not available for the whole area of interest, so that the following situations were evident for certain municipalities:

- partial or complete lack of data on buildings
- different formats and ways of layer presentation
- incompleteness of the data, including census numbers, inability to connect geometry with attribute data and other problems.

Also, no data on the district heating distribution network was available for a number of municipalities.

Because of these reasons different approaches have been used to compensate these shortcomings. For most of the municipality, data on buildings is downloaded from the Internet (Open Source Maps) and data are formatted to a unique format for applying a uniform GIS processing model. For some municipalities, it was not possible to determine the demands of the non-residential sector due to lack of data on buildings. Most of the data is taken from the cadaster databases resulting in the lack of many buildings and the unreliability of their areas due to timeliness.

After determining the key indicators of the situation with the energy needs, it is followed their geovisualization and creating of intelligent maps with different thematic levels (by point symbols, by municipalities, by grids 1x1 km and 100x100 m resolution). This implied the creation of several maps of heating demands (Figure 8), cooling demands (Figure 9), the potential of waste heat, the potential of renewable energy sources including solar energy, biomass, geothermal energy, and finally maps of existing district heating systems and the potential of their expansion. For the

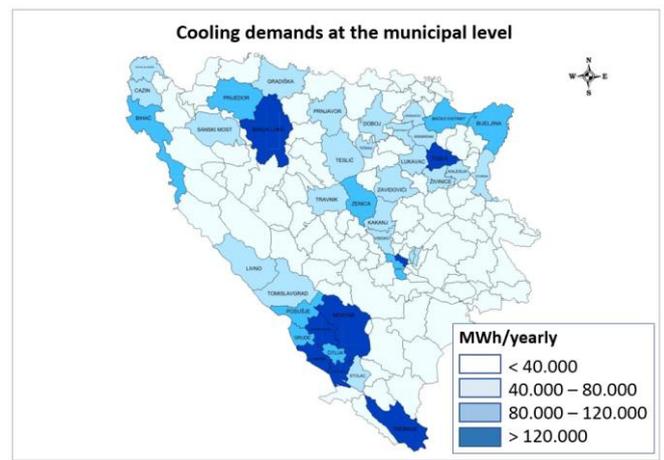


Figure 8. Map of cooling demands distribution by municipalities purposes of map creation, the appropriate models and designing procedures for each type of map were prepared followed by an appropriate set of data from the database.

For the purposes of mapping "top down", ie the determination of priority areas for further analysis, statistical data was collected at the level of individual municipalities. These data included the population, the total areas of all types of building and the climatic zones. Using these data and calculating the specific demand for heating and cooling for particular types of objects, demand for heating and cooling of each municipality was determined. These data were then distributed spatially by the CORINE map. Sources of waste heat and the potential amounts of energy they can deliver to the system are determined through collected data on greenhouse gas emissions according to the proposed methodology. The "top down" methodology used in this case was fully applied and no significant deviations were noted.

However, due to the lack of some key data needed to conduct „bottom-up“ analysis, the initial methodology had to be partially adapted. As cadastral data was not available for all the observed municipalities, and for some missing significant amounts of data, the application of Open Source Maps data was suggested, but this source was again insufficiently detailed. Data on the altitudes and the building purpose were also not available in a large part of the

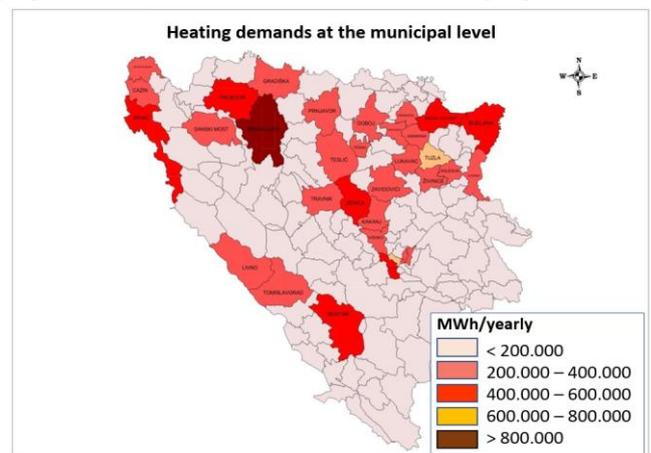


Figure 9. Map of heating demands distribution by municipalities

municipalities because they were either not public or not at all. Analyzing literature, it looks that are all common problems of similar projects. [24] [25]

In order to achieve the required goals, data at the level of census circles (which areas are mainly below 100X100 m) is applied, which gave results of comparable quality.

The proposed methodology could not be fully realized in this case due to the lack of key data, but the proposed alternatives are acceptable and have achieved satisfactory quality and level of detail.

#### V. CONCLUSION

In order to ensure the successful implementation of a comprehensive assessment of the potential for high efficiency cogeneration and efficient centralized heating and cooling, it is necessary to know the spatial distribution of their demand as well as potential sources of waste heat. Since the process of such mapping is highly dependent on the availability of data, it is initially proposed two step approach including methods of state mapping by the "top to bottom" and local mapping by "bottom up". The purpose of such an approach is, on the one hand, to create a coherent map with distributions of demand for heating and cooling and potential sources of waste heat for the whole observed area with a grid in resolution 1x1 km. On other hand, local mapping of selected areas gives a more detailed view of the state by applying a grid in resolution 100x100 meters [5].

For the final procedure applied for the assessment of the potential for the application of high-efficiency cogeneration and efficient district heating and cooling, they are suggested alternatives which were proved as acceptable. The main reason for the deviation from the initial methodology is the lack of data on which the methodology depends. This applies in particular to the data required in the "bottom up" mapping, ie to the current cadastral data on buildings. In order to conduct the process with an adequate level of detail, it is used census circuit data which generally have a resolution of less than 100X100 meters and are therefore applicable for the described procedure. The obtained results are in line with what the default methodology prescribes.

Randomized control of the obtained results indicate that applied data sets and appropriate geospatial algorithms satisfy the required level of accuracy and reliability implied by this methodology.

#### REFERENCES

- [1] "Energy Efficiency Directive 2012/27/EU", Available: <https://eur-lex.europa.eu/eli/dir/2012/27/oj> [Accessed: 7-Jan-2018]
- [2] Final Report on Assessment of the Potential for the Application of High-efficiency Cogeneration and Efficient District Heating and Cooling, Ceteor Sarajevo, 2018
- [3] "Stratego." [Online]. Available: <http://stratego-project.eu/>. [Accessed: 30-Aug-2017]
- [4] "Heat Roadmap Evrope." [Online]. Available: <http://www.heatroadmap.eu/>. [Accessed: 15-Jun-2017]
- [5] N. Duic, T. Puksec, and T. Novosel, "Methodology for assessment of the potential for the application of high-efficiency cogeneration and efficient district heating and cooling - Development of state and local GIS maps", Sarajevo, 2018
- [6] "CORINE Land Cover — Copernicus Land Monitoring Service." [Online]. Available: <https://land.copernicus.eu/pan-evropean/corine-land-cover>. [Accessed: 10-Jan-2018]
- [7] A. Karabegovic, M. Ponjavic, "Integration and Interoperability of Spatial Data in Spatial Decision Support System Environment", MIPRO IEEE Croatia Conference, Opatija, Croatia, 2010
- [8] P. Longley, M. Goodchild, D. Maguire, D. Rhind, „Geographic Information Systems and Science“, John Wiley & Sons, 2002
- [9] M. Ponjavic, "Basics of Geoinformation", Faculty of Civil Engineering, University of Sarajevo, 2011
- [10] M. de Smith, M. Goodchild, P. Longley, "Geospatial Analysis – A Comprehensive Guide to Principles Techniques and Software Tools, <https://www.spatialanalysisonline.com>, 2018
- [11] R. Haining, "Spatial Data Analysis: Theory and Practice", Cambridge University Press, 2003
- [12] D. Arnautović-Aksić et al., Typology of residential buildings of Bosnia and Herzegovina, GIZ Sarajevo, 2016
- [13] M. Nisandžić et al., Typology of Public Buildings in Bosnia and Herzegovina, UNDP Sarajevo, 2017
- [14] Feasibility Study of Animal By-products and Animal Waste Management in Bosnia and Herzegovina, EPRD Office for Economic Policy and Regional Development Ltd. Poland, 2018
- [15] U. Persson, "Quantifying the Excess Heat Available for District Heating in Evrope," p. 17, 2015
- [16] U. Persson, B. Moller, and S. Werner, "Heat Roadmap Evrope: Identifying strategic heat synergy regions," Energy Policy, vol. 74, pp. 663–681, 2014
- [17] A. Karabegovic and M. Ponjavic, "Geoportal as decision support system with spatial data warehouse," 2012 Federated Conference on Computer Science and Information Systems (FedCSIS), Wroclaw, 2012, pp. 915-918
- [18] M. Ponjavic, and A. Karabegovic, "Location Intelligence Systems and Data Integration for Airport Capacities Planning", Computers. 2019; 8(1):13, <https://doi.org/10.3390/computers8010013>
- [19] V. Somogyi, V. Sebestyén, and E. Domokos, "Assessment of wastewater heat potential for district heating in Hungary", Energy, vol 163, 2018, pp. 712-721, ISSN 0360-5442, <https://doi.org/10.1016/j.energy.2018.07.157>.
- [20] A. Dénarié, M. Muscherà, M. Calderoni, and M. Motta, "Industrial excess heat recovery in district heating: Data assessment methodology and application to a real case study in Milano", Italy, Energy, vol 166, 2019, pp. 170-182, ISSN 0360-5442, <https://doi.org/10.1016/j.energy.2018.09.153>.
- [21] R. Buffat, and M. Raubal, "Spatio-temporal potential of a biogenic micro CHP swarm in Switzerland", Renewable and Sustainable Energy Reviews, vol 103, 2019, pp. 443-454, ISSN 1364-0321, <https://doi.org/10.1016/j.rser.2018.12.038>.
- [22] M. Segal, "Location always matters: how to improve performance of dynamic networks?", 2016 Federated Conference on Computer Science and Information Systems, M. Ganzha, L. Maciaszek, M. Paprzycki (eds). ACSIS, Vol. 8, pp. 5–5, 2016. <http://dx.doi.org/10.15439/2016F596>
- [23] B. Prokop, J. Owsiński, K. Sep, P. Sapiecha, "Solving the k -Centre Problem as a method for supporting the Park and Ride facilities location decision", 2016 Federated Conference on Computer Science and Information Systems, M. Ganzha, L. Maciaszek, M. Paprzycki (eds). ACSIS, Vol. 8, pages 1223–1228, 2016. <http://dx.doi.org/10.15439/2016F300>
- [24] E. Ziemba, "The ICT Adoption in Government Units in the Context of the Sustainable Information Society", Federated Conference on Computer Science and Information Systems, M. Ganzha, L. Maciaszek, M. Paprzycki (eds). ACSIS, Vol. 15, pp. 725–733, 2018. <http://dx.doi.org/10.15439/2018F116>
- [25] P. Ziemba, J. Wątróbski, A. Karczmarczyk, J. Jankowski, W. Wolski, "Integrated Approach to e-Commerce Websites Evaluation with the Use of Surveys and Eye Tracking Based Experiments", 2017 Federated Conference on Computer Science and Information Systems, M. Ganzha, L. Maciaszek, M. Paprzycki (eds). ACSIS, Vol. 11, pp 1019–1030, 2017. <http://dx.doi.org/10.15439/2017F320>

# MOPS – A feasibility study for working with GPS and sensor data in a medical context

Christof Meigen

Leipzig Research Center for  
Civilization Diseases – LIFE Child  
Email: cmeigen@life.uni-leipzig.de

Mandy Vogel

Center for Pediatric Research Leipzig, University  
Hospital for Children and Adolescents  
Email: vogel@medizin.uni-leipzig.de

Jan Bumberger

Helmholtz-Centre for Environmental  
Research - UfZ  
Email: jan.bumberger@ufz.de

**Abstract**—New kinds of data collection like GPS-tracking, wearable sensors and mobile apps impose both technical and privacy challenges for medical research. In the MOPS study (*Machbarkeitsstudie für Ortsbezogene Parameter und Sensordaten* – feasibility study for geocoded parameters and sensor data) we provided 10 participants with a newly developed app and sensors for various physical and environmental parameters. We want to explore the feasibility of the recently established Medical Research Platform (MRP) of the Medical Faculty of the University of Leipzig and similar platforms for this kind of data collection and processing.

After briefly describing the Medical Research Platform we report on the technical set-up of the MOPS project in this setting and first practical experiences.

## I. INTRODUCTION

**T**ECHNICAL advances in the last decade – especially the ubiquity of smartphones – have made new kinds of data collection feasible for research. Sensors for many physical parameters are now comfortably wearable. Public facilities and initiatives for Open Data make more and more datasets publicly available, and it has become easy to – for example – link individual GPS data to public land use or noise maps. Software to work with this data is also freely available.

Medical research in particular is rooted in a tradition with strong focus on ethical and privacy consideration[2], and on long-term reproducibility on the results from raw data. That may sound trivial at first, but in practice it means getting explicit approval from an ethics board for each specific data collection, and storing all your raw data and analysis scripts for at least 10 years according to Good Clinical Practice.

### A. Privacy Issues

The collection of vast amounts of data about an individual raises serious questions about privacy. Long gone are the days that just using a pseudonym for each participant was viewed as a sufficient protection against re-identification. GPS data reveals your home and work address, answers to questionnaires might be matched against social media profiles and high-resolution sensor data from physical parameters contain highly specific individual patterns. Ever-present timestamps can be used to identify events and may themselves contain sensitive information.

An innocent looking data point like `{lat: 51.30175, long: 12.3775013, timestamp: "2019-05-20`

`17:32:12"}` could already be proof that the person this record refers to is an alcoholic (it's the time and place of an AA meeting).

For research, German law requires the “separate storage” of identifying data and research parameters (§27(3) BDSG). While this already means you should be using pseudonyms and store the names and contact information of study participants in a different database (or at least a different database table), it has been interpreted in the past as a requirement to also store research parameters with higher risk of re-identification like MRI data or genetic data separated *from each other* using different pseudonyms and a separate system (ideally managed by a trusted third party) to store the connection between these pseudonyms[6]. As discussed, tracking data from sensors certainly falls into the same category and should be treated accordingly.

### B. Reproducibility

In a setting where research datasets are basically CSV files with one row per participant and one column per variable, and the – previously committed to – analysis plan consists of a few well-understood statistical tests, reproducibility can be achieved by archiving a few data files in a text-based format. The description in the publication is often sufficient to redo the calculations.

Nowadays, however, data is often requested on the fly from various data sources, and sophisticated software packages (with many dependencies on other packages) are used to process the data.

Reproducing results – especially many years later – requires archiving not only data in various formats, but also scripts to automatically obtain the published results from the data. For this, archiving the exact software environment used for analysis is often necessary due to changes in packages, incompatibility of new versions or deprecation of features. This is also in line with more recent general requirements for scientific data management like the FAIR data principles[5]. It is however at odds with workflows that rely on online services (which might become unavailable or return different results), specific versions of proprietary software (whose licence key might expire) and in general complex systems of interacting parts.

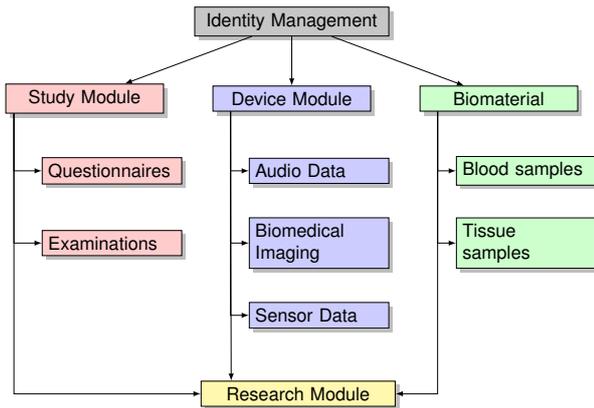


Fig. 1. Main Modules of the Medical Research Platform

### C. Scope of the MOPS study

With the MOPS study, we build on previous considerations on combining sensor data and publicly available data with classical approaches from epidemiological research laid out by Kirsten et. al. [1].

We use a small sample set ( $n=10$ ) to test the feasibility of the recently established Medical Research Platform (MRP) of the Medical Faculty of the University of Leipzig for this kind of study.

## II. THE MEDICAL RESEARCH PLATFORM

### A. Overview of the Modules

The study management software REDCap[3] has been used extensively at the Medical Faculty of the University of Leipzig since 2013 to conduct studies, especially those that are not part of a drug approval process. Electronic case report forms (eCRF's) are easy to set up and the system provides excellent support for various data management tasks.

In light of the new General Data Protection Regulation (GDPR) the use of REDCap was re-evaluated in 2018 and confirmed as a platform for future research projects – but it was amended by a separate ID- and Consent-management system to store identifying information and connection between pseudonyms separate from each other and from the REDCap system. Additionally, nextcloud-based file archives have been set up to separately store data from devices and to archive research datasets, and the LabCollector LIMS has been installed to track biomaterial samples.

A data protection concept was drafted to define the various modules of the Medical Research Platform (see Figure 1 – the Biomaterial module is not used in the MOPS study) and to describe what data (under which pseudonyms) is stored in each module, what the interfaces between the modules are and especially how and when the re-pseudonymisation is performed and how access rights are granted.

### B. ID Management

The ID Management solution LEIM was developed in coordination with the Data Integration Center of the Uni-



Fig. 2. The main screen of the MOPS app

versity of Leipzig Medical Center to serve as a flexible model implementation for a *Patient Identifier Cross-reference Manager*, a *Consent Management System* (accessible only via API) and a separate Web application for Contact Management. It's main focus is the smooth integration with the other components of the Medical Research Platform and an ongoing adaptation to the concepts and interfaces established by the Data Integration Center as part of the German Medical Informatics Initiative (SMITH). Currently LEIM does not use a so-called *PID-Generator* like the *Mainzliste*[4], but instead the pseudonymisation service generates a random Contact ID (*KID*) for use in the contact management and links it internally to the (also randomly generated) *PID*.

## III. THE MOPS STUDY

### A. Overview

In the MOPS project we equip participants with an app for GPS and mood tracking (see Figure 2), a wristband (bodymonitor.de) to collect physiological parameters like skin conductivity and temperature, and a sensor for environmental parameters like air humidity and temperature (FreeTec, Model NC-7004-675). Participants are asked to wear these devices for at least 24 hours, preferably longer.

It is a feasibility study to establish technical and organisational processes and to investigate the usability of the Medical Research Platform for this kind of data collection and processing.

Data	Source	Nr of records
GPS location data	App	286 805
Stress & Movement	BodyMonitor	4416 083
Mood	App (manual)	270
Transportation	App (manual)	196
Temperature/Humidity	FreeTec Sensor	2 504
Questionnaires	Interview	20

TABLE I  
DATA COLLECTED IN THE MOPS STUDY

### B. Organisatorical preliminaries

In order to conduct the study we submitted a study protocol to the Ethics committee. The study protocol did not only include the precise description of collected data, the aim of the study, description of recruitment process but also a data protection impact assessment detailing the risks for the participants in case of data leaks.

### C. Collected data

We collected data from 10 participants with the wristband, MOPS-App and questionnaires. Two of the participant used an additional sensor for surrounding temperature and humidity. An overview of the collected data points can be seen in Table I.

After collecting the data, participants were interviewed for their experiences. Most people found the wristband and the notifications of the app annoying or slightly annoying, while 7 out of 10 participants used Google Services to improve the accuracy of the location information, thereby transferring all the location data collected in the MOPS study also to Google.

Data cleaning and matching the various time-related data is still ongoing, but the overall functioning of the data collection was verified during a piloting phase (see Figure 3).

## IV. PRACTICAL EXPERIENCES WITH THE MRP

### A. LEIM for ID and consent management

For contact management, ID management and consent management a solution called LEIM (*Leipziger Einwilligungs-und Identitätsmanagement* – Leipzig Consent and identity management) had been developed based on the experience with similar tools especially at the German Center for Neurodegenerative Diseases.

The process of setting up a new study, defining the Informed Consent form and the types of Pseudonyms used is done through a simple web-based interface. Pseudonym types are defined by a simple declaration of fixed parts (most often prefixes) and random parts. The pseudonymisation service makes sure that random parts are created by a cryptographically secure pseudorandom number generator and that pseudonyms are unique within a study.

We defined the following pseudonyms: a Study Identification Code (*SIC*), an ID for the MOPS-App, an ID for the wristband, an ID for the sensor, and a pseudonym for the research dataset (*PSN*). In addition, each participant gets assigned a leading person identifier *PID* in the pseudonymisation service

of the Identity Management and an contact identifier *KID* in the contact management part of the Identity Management.

After that, a study- and role-specific token was created which allows requests to the pseudonymisation service of LEIM from other modules in order to map pseudonyms and to request consent status. Access to identifying information (names, addresses) is not possible through these requests.

As a last step, the project ID in the REDCap system has to be entered to allow easy referrals from contact management to data entry forms in REDCap (which requires the appropriate role for the user in both systems).

### B. REDCap for study data

In the MOPS study we collect only basic sociodemographic data – age and gender – as well as body height and weight in the study module. The pseudonyms for the devices (which are entered in the app, and stored on the the wristband and attached to the sensor) are stored only in LEIM, not in REDCap. Data entry and data export workes flawlessly in REDCap, as expected.

### C. Nextcloud for device data and research data archive

Nextcloud is an open-source, self-hosted file share and collaboration platform. It stores the uploaded files on the host file system, but provides checksums and versioning for all files, fine grained access control, access through a web interface as well as mounting it as a network drive (through WebDAV) and even a client for local synchronisation (not currently used in the Medical Research Platform). Using nextcloud is absolutely straightforward for any computer user.

## V. PRACTICAL EXPERIENCES WITH TECHNOLOGIES USED IN MOPS

Apart from the technical platform already provided by the Medical Research Platform we also explored various tools specific to the collection of sensor data.

### A. Flutter for App development

While there is certainly no shortage of mobile apps, especially with respect to monitoring anything health/fitness related, we wanted to explore how difficult it is to create a simple app where we could completely control the data collection and data transfer process.

Having a background in web development and scripting languages, we evaluated mobile frameworks that provided a familiar development workflow, especially having no (re-)compile times, a simple dynamic language and an accessible set of simple cross-platform widgets. We chose Flutter (flutter.io) and were able to create an app with GPS tracking, uploading, notifications to regularly enter the current mood status and mode of transportation within 3 days of work, resulting in little more than 500 lines of code.

The only compromise we made was not using GPS tracking in the background (i.e. when the app is not running), as this is currently only available as a commercial plugin.

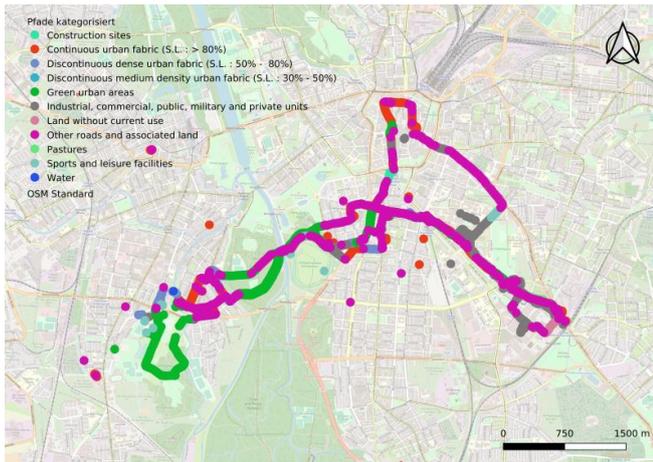


Fig. 3. A path recorded by the MOPS app, categorized by land use using the GeoEtiology PostGIS database

### B. PostGIS and QGIS for Geodata

As part of the long-term GeoEtiology project a PostgreSQL/PostGIS database was set up containing geocoded information about the Leipzig region from various sources. This includes noise maps, the road network, information about land use, places of interest (restaurants, schools), trees etc. Information about the social characterisation of a large number of addresses was purchased from the *SINUS Markt- und Sozialforschung GmbH*.

For the MOPS study, the database acted as an (rather large) part of the data processing pipeline in the device module. We loaded the data temporarily into the database, ran some queries involving the various spatial information stored there, and exported the aggregated results, see for example Figure 3 for a dataset from piloting where a path was categorized according to land use.

Although we repeatedly encountered queries that required some restructuring, additional indices and usage of materialized views to perform well, the overall experience with the GeoEtiology PostGIS database has been splendid.

### C. Guix for reproducible environments

The Medical Research Platform currently provides no special environment for scientific computing, instead, researchers are encouraged to ensure the reproducibility of their postprocessing and analysis scripts themselves by defining the required software environment alongside the scripts.

There are various lightweight approaches for different programming languages – virtualenv for Python, packrat for R, the Manifest file for Julia – but having whole virtual environments of all the software tools used has in the past been limited to container-based solutions like Docker.

We instead defined Guix environments[7] using the concept of channels, where specific versions of all used software (including R and Python packages, but also R and Python themselves) can be defined. All dependencies down to the operating system kernel are then tracked and set up. The channel definitions are simply checked into the version control

system alongside the scripts. Switching between environments is instantaneous, and reproducing environments on a different computer might require downloading packages but is otherwise guaranteed to reproduce the exact same results. While we used Guix environment for data processing scripts, we do currently not run the GeoEtiology database in such an environment.

## VI. CONCLUSION

Conducting the MOPS study on the Medical Research Platform is an interesting experience. The usage of different pseudonyms for different kinds of data and repseudonymisation for research datasets with the help of a separate ID management might seem cumbersome and overly cautious at first, but in practice it works well using the API of the LEIM pseudonymisation service. Generally, privacy concerns should be taken seriously and addressed at every step of the data processing.

Consequently working within Guix environments to ensure reproducibility seems doable in daily practice at least for the currently rather limited set of postprocessing and analysis scripts.

We hope to soon report on analysis results from the dataset collected in the MOPS study.

## ACKNOWLEDGMENT

The authors wish to thank Prof. Dr. Antje Körner, PI of the GeoEtiology project, for her support in using the GeoEtiology infrastructure. The activities were co-financed by the research project *Smart Sensor-based Digital Ecosystem Services* (S2DES, 2016-2020), funded by the European Social Fund (ESF; Grant Agreement No. 100269858)

## REFERENCES

- [1] Kirsten T., Bumberger J. et al. (2017), “Research in Progress on integrating Health and Environmental Data in Epidemiological Studies,” In *Abramowicz W., Alt R., Franczyk B. (eds) Business Information Systems Workshops. BIS 2016. Lecture Notes in Business Information Processing, vol 263. Springer, Cham*, doi: 10.1007/978-3-319-52464-1\_32
- [2] World Medical Association. (2001), “World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects”, *Bulletin of the World Health Organization*, 79 (4), 373 – 374. World Health Organization. <http://www.who.int/iris/handle/10665/268312>
- [3] Paul A. Harris, Robert Taylor, Robert Thielke, Jonathon Payne, Nathaniel Gonzalez, Jose G. Conde (2009), “Research electronic data capture (REDCap) – A metadata-driven methodology and workflow process for providing translational research informatics support”, *J Biomed Inform.* 2009 Apr;42(2):377-81, doi: 10.1016/j.jbi.2008.08.010
- [4] M Lablans, A Borg, F Ückert (2015), “A RESTful interface to pseudonymization services in modern web applications”, *BMC Med Inform Decis Mak.* 2015 Feb 7;15:2. doi: 10.1186/s12911-014-0123-5.
- [5] Wilkinson MD, Dumontier M, Jan Aalbersberg I, et al. (2016), “The FAIR Guiding Principles for scientific data management and stewardship”, *Sci Data.* 2016(3), article number 160018, doi:10.1038/sdata.2016.18
- [6] Pommerening K, Drepper J, Helbing K, Ganslandt T (2014), “Leitfaden zum Datenschutz in medizinischen Forschungsprojekten – Generische Lösungen der TMF 2.0”, ISBN: 978-3-95466-123-7
- [7] Courtès L, Wurmus R (2015): “Reproducible and User-Controlled Software Environments in HPC with Guix” In *Euro-Par 2015: Parallel Processing Workshops*, 579–591, Springer International Publishing, Cham, doi: 10.1007/978-3-319-27308-2\_47

# 14<sup>th</sup> Conference on Information Systems Management

**T**HIS event constitutes a forum for the exchange of ideas for practitioners and theorists working in the broad area of information systems management in organizations. The conference invites papers coming from two complimentary directions: management of information systems in an organization, and uses of information systems to empower managers. The conference is interested in all aspects of planning, organizing, resourcing, coordinating, controlling and leading the management function to ensure a smooth operation of information systems in an organization. Moreover, the papers that discuss the uses of information systems and information technology to automate or otherwise facilitate the management function are specifically welcome.

## TOPICS

- Management of Information Systems in an Organization:
  - Modern IT project management methods
  - User-oriented project management methods
  - Business Process Management in project management
  - Managing global systems
  - Influence of Enterprise Architecture on management
  - Effectiveness of information systems
  - Efficiency of information systems
  - Security of information systems
  - Privacy consideration of information systems
  - Mobile digital platforms for information systems management
  - Cloud computing for information systems management
- Uses of Information Systems to Empower Managers
  - Achieving alignment of business and information technology
  - Assessing business value of information systems
  - Risk factors in information systems projects
  - IT governance
  - Sourcing, selecting and delivering information systems
  - Planning and organizing information systems
  - Staffing information systems
  - Coordinating information systems
  - Controlling and monitoring information systems
  - Formation of business policies for information systems
  - Portfolio management,
  - CIO and information systems management roles
- Information Systems for Sustainability
  - Sustainable business models, financial sustainability, sustainable marketing
  - Qualitative and quantitative approaches to digital sustainability
  - Decision support methods for sustainable management

## EVENT CHAIRS

- **Arogyaswami, Bernard**, Le Moyne University, USA
- **Chmielarz, Witold**, University of Warsaw, Poland
- **Jankowski, Jarosław**, West Pomeranian University of Technology in Szczecin, Poland
- **Karagiannis, Dimitris**, University of Vienna, Austria
- **Kisielnicki, Jerzy**, University of Warsaw, Poland
- **Ziemia, Ewa**, University of Economics in Katowice, Poland

## PROGRAM COMMITTEE

- **Aguillar, Daniel**, Instituto de Pesquisas Tecnológicas de São Paulo, Brazil
- **Bontchev, Boyan**, Sofia University St Kliment Ohridski, Bulgaria
- **Cano, Alberto**, Virginia Commonwealth University, United States
- **Cingula, Domagoj**, Economic and Social Development Conference, Croatia
- **Czarnacka-Chrobot, Beata**, Warsaw School of Economics, Poland
- **Damasevicius, Robertas**, Kaunas University of Technology, Lithuania
- **Deshwal, Pankaj**, Netaji Subash University of Technology, India
- **Duan, Yanqing**, University of Bedfordshire, United Kingdom
- **Eisenhardt, Monika**, University of Economics in Katowice, Poland, Poland
- **El Emery, Ibrahim**, King Abdulaziz University, Saudi Arabia
- **Espinosa, Susana de Juana**, University of Alicante, Spain
- **Fantinato, Marcelo**, University of Sao Paulo, Brazil
- **Gabryelczyk, Renata**, University of Warsaw, Poland
- **Gawel, Aleksandra**, Poznan University of Economics and Business, Poland
- **Geri, Nitza**, The Open University of Israel, Israel

- **Halawi, Leila**, Embry-Riddle Aeronautical University, United States
- **Kania, Krzysztof**, University of Economics in Katowice, Poland
- **Kobyliński, Andrzej**, Warsaw School of Economics, Poland
- **Leyh, Christian**, University of Technology, Dresden, Germany
- **Michalik, Krzysztof**, University of Economics in Katowice, Poland
- **Mullins, Roisin**, University of Wales Trinity Saint David, United Kingdom
- **Muszyńska, Karolina**, University of Szczecin, Poland
- **Nuninger, Walter**, Polytech'Lille, Université de Lille, France
- **Rizun, Nina**, Faculty of Management and Economy, Gdansk University of Technology, Poland
- **Rozevskis, Uldis**, University of Latvia, Latvia
- **Schroeder, Marcin Jan**, Akita International University, Japan
- **Sobczak, Andrzej**, Warsaw School of Economics, Poland
- **Swacha, Jakub**, University of Szczecin, Poland
- **Symeonidis, Symeon**, Democritus University of Thrace, Greece
- **Szczerbicki, Edward**, University of Newcastle, Australia
- **Szumski, Oskar**, University of Warsaw, Poland
- **Travica, Bob**, University of Manitoba, Canada
- **Wątróbski, Jarosław**, University of Szczecin, Poland
- **Wielki, Janusz**, Opole University of Technology, Poland
- **Zaitsev, Dmitry**, Vistula University, Poland

# Exploring Determinants of M-Government Services: A Study from the Citizens' Perspective in Saudi Arabia

Mohammed Alonazi  
University of Sussex  
Informatics Department  
Brighton, UK  
Email: M.Alonazi@sussex.ac.uk

Natalia Beloff  
University of Sussex  
Informatics Department  
Brighton, UK  
Email: N.Beloff@sussex.ac.uk

Martin White  
University of Sussex  
Informatics Department  
Brighton, UK  
Email: M.White@sussex.ac.uk

**Abstract**— The government of Saudi Arabia has adopted M-Government for the effective delivery of services. One advantage that it offers is unique opportunities for real-time and personalized access to government information and services. However, a low adoption rate of m-Government services by citizens is a common problem in Arab countries, including Saudi Arabia, despite the best efforts of the Saudi government. Therefore, this paper explores the determinants of citizens' intention to adopt and use m-Government services, in order to increase the adoption rate. This study was based on the Mobile Government Adoption and Utilization Model (MGAUM) that was developed for the purpose. Data was collected, and the final sample consisted of 1,286 valid responses. The descriptive analysis presented in this paper indicates that all the proposed factors in our MGAUM model were statistically significant in influencing citizens' intention to adopt and use m-Government services.

## I. INTRODUCTION

Governments from across the world have digitized their services to citizens through mobile technologies and the Internet, which has arguably improved communication between citizens and their governments, to provide better access to services and information, as well as improving government accountability, transparency and public governance [1], [2]. In this study, m-Government is defined as the use of mobile technology to deliver and improve e-Government services and information to citizens, commercial organisations and all government agencies. Previous studies have either regarded m-Government as separate to e-Government or as an extension or replacement of it [3], [4]. As it offers the public a valuable extra means to access services and information, it can be considered an advance in government service delivery [5]. Although e-Government and m-Government work on the same principle, the latter is distinguished by features that are particular to it: Citizens can access the network from anywhere and at any time [6]; and can instantly receive messages from government service providers on their mobiles [7]. The mobile phone has recently become the primary way people communicate and has thus arguably become a part of everyday life for many people [8]. Therefore, being able to access government services via mobile devices might be the

best route for citizens. Accessing government services and information on their phones means citizens neither have to visit the service provider in person nor go home to use their computer in order to do this [7], [9]. Access to the Internet may depend on economic factors, i.e. the extent of Internet access in a particular country, and how many citizens have access to computers providing mobile services can overcome these limitations [5].

An m-Government system provided by wireless technology, will give citizens opportunities for personalized access to government services and information in real time [10], [6]. This is especially beneficial for users in remote areas as m-Government services have the advantages of being affordable and easily and immediately accessible. Further, a relatively low level of digital literacy is required to operate them successfully [11], [6]. Given these features, the adoption of an m-Government system benefits both citizens and governments.

The Mobile Government Adoption and Utilization Model (MGAUM) has been developed as a framework from which to analyse factors affecting adoption and use of m-Government services [12]. This study aims to investigate and understand Saudi citizens' perceptions towards the adoption and utilization of m-government services in developing countries, particularly Saudi Arabia, in order to increase the adoption rate of m-government services.

## II. THE RESEARCH METHODOLOGY

The study was conducted in Saudi Arabia; and the questionnaire was distributed to adult Saudi citizens. The total valid responses constituted a sample of 1,286 Saudi citizens. SPSS was used to analyse the survey study data. The survey questionnaire was developed and modified from items used in previous research into both e-Government and m-Government. All items were measured with a 5-point Likert scale.

### A. Reliability and validity of the study

To be considered reliable, a research instrument needs to produce similar results if used in comparable conditions and be relatively free of errors [13]. As part of the pilot study for

this research, Cronbach's Alpha was chosen to assess the reliability of the questionnaire, and establish the internal consistency of the constructs used. Table 1 indicates the Cronbach's Alpha results for the complete research instrument, and demonstrates that the reliability of each of the constructs (independent and dependent variables) lies within the range of what is thought of as acceptable in academic research.

Face and content validity were chosen to be investigated rather than construct or criterion validity. The face validity of each item was ascertained in the pilot study to ensure that the model's factors measured what they were intended to measure; and items deemed to lack sufficient clarity, unambiguity or relevance were deleted or revised accordingly.

TABLE 1: INTERNAL CONSISTENCY OF THE STUDY SURVEY INSTRUMENT

Measured Variable	No of Items	Cronbach's Alpha
Perceived Usefulness (PU)	7	.898
Perceived Ease of Use (PEOU)	4	.862
Social Influence (SI)	3	.779
Perceived Compatibility (PCOM)	2	.868
Perceived Trust (PT)	7	.715
Culture (CULT)	5	.616
Awareness (AW)	4	.842
Perceived Mobility (PM)	3	.776
Citizens Service Quality (CSQ)	8	.920
System Quality (SQ)	7	.745
Intention to use (ITU)	4	.894

### III. RESEARCH FINDINGS AND DISCUSSION

This section provides an overview of respondents' demographic characteristics; and a descriptive analysis for each factor proposed in the MGAUM is given in order to explain their impact on citizens' intention to adopt and use m-Government services in Saudi Arabia.

#### A. Respondents' demographic data

In the final sample 813 participants were male (63.2%) and 473 were female (36.8%). The highest percentage of participants was in the 18–30 age group, the largest number held a bachelor's degree and over half the participants (46.7%) were government employees. All participants had Smartphone devices, with a large majority using mobiles and the Internet in daily life (96.4% and 93.3% respectively). Approximately three-quarters of the respondents (76.7%) have some knowledge about m-Government services in Saudi Arabia, whereas 22.9% had no knowledge. Moreover, the majority of participants (90.5%) already used m-Government services, but 9.5% had never used them. Participants who had already used m-Government services were asked to rate their general experience: 40.3% were very satisfied, 44.3% were satisfied to some extent, with only 5.9% not satisfied. Furthermore, 22.6% reported that the

requirements of the intended m-Government services were not clear and 26.6% reported that the system quality of m-Government services was not good.

#### B. The Measures of central tendency and Likert Scale

A central tendency sums up an entire set of differing values, so the mean, median or mode is used according to what is most appropriate for the specific conditions being described. The mean is the most common measure of central tendency and was used in this study [14].

A Likert scale, was chosen as the main instrument in this study's questionnaire, as the simplest and most practical way to measure strength of opinion; and a review of the literature shows that is most commonly and successfully employed in IS research [14,15].

Weighted averages were calculated for the Likert scales, from Strongly Agree=1 to Strongly Disagree=5, (see Table 2) so that the tendency of the composite scores could be ascertained. The numbers entered into SPSS represent 'weight' and the weighted averages for the scale needs to be calculated to understand means. The results can be interpreted to show how influential (or not) each factor is [16].

TABLE 2: WEIGHTED AVERAGES FOR 5-POINT LIKERT SCALES

Weighted average	Result	Result Interpretation
1 – 1.79	Strongly agree	Very influential
1.80 – 2.59	Agree	Influential
2.60 – 3.39	Neutral	Neutral or do not know
3.40 – 4.19	Disagree/	Uninfluential
4.20 – 5	Strongly disagree	Very uninfluential

#### C. Descriptive analysis of data

The data collected was analysed with reference to each of the constructs in the MGAUM [12]. Participants' attitudes, intentions and behaviour towards adopting and using Saudi m-government services were explored by means of responses to statements for which Likert scores could be calculated.

TABLE 3: RESULT FOR ALL FACTORS

Factor	Items	Mean	S.D.	Result interpretation
PU	7	1.3900	.49876	Very influential
PEOU	4	1.5733	.61942	Very influential
CULT	5	1.9997	.69224	Influential
PT	7	2.3586	.62882	Influential
SI	7	1.6516	.62692	Very influential
PCOM	2	1.5638	.66208	Very influential
AW	4	2.1763	.87356	Influential
CSQ	8	2.0852	.69808	Influential
SQ	7	1.8891	.54653	Influential
PM	3	1.5625	.57052	Very influential

Table 3 summarizes the results of the descriptive analysis with an interpretation of all results.

**Perceived Usefulness (PU):** It was important to determine what users' perceptions were about the usefulness of m-

Government, the advantages that they would gain from using m-Government services in terms of saving time, effort and money and how this influenced their behavioural intention.

Most (93.9%) agreed that using mobile government services would be useful. Participants believed that using them would enable government transactions to be effected more quickly as well as saving them time, money and effort (94.4%) and enable them to perform transactions from any location (96%); 89.6% believed that using m-Government services would make communication between a government agency and citizens easier through text message, applications and e-mail; and over 94.2% agreed that using mobile government services would remind them of important dates in order to conduct or receive government transactions in sufficient time or at the right time. Using features on mobile devices that are not found on the website, such as reminders, location and camera were thought to add value to services.

The factor was interpreted as significantly influential on citizens' intention to adopt and use m-Government services; and that use of m-Government services would increase when users perceive their benefits. This indicated that governments should take into consideration the requirements and needs of users to be met before implementing any services. The findings of this study are consistent with previous empirical investigations of factors that might affect the adoption of m-Government services in Egypt [17], and Taiwan [18].

**Perceived Ease of Use (PEOU):** The vast majority of participants (93.7%) agreed that learning to use mobile government services would be easy; and 92% believed that interactions with m-Government would be clear and understandable; 86.2% believed that using m-Government services required little skill and effort.

The total score for the PEOU factor was 1.5733 (see Table3), which indicated that PEOU is very influential on users' intention to adopt and use M-Government services. The result suggests that the user experience is the first step in adoption; and if a user finds m-Government services easy to use and that it saves time and effort, this impacts positively on behavioural intention to adopt and use them. If services are easy to use, and people do not have to rely on asking for help from another person to use the application, the number of users will increase. Thus, PU and PEOU are essential factors in the MGAUM, and any theoretical framework which seeks to analyse intention, adoption and use of m-government in the Saudi context or similar contexts in developing countries. The findings are with line with previous studies which found that PEOU is an important factor in determining intention to use [18], [19].

**Culture (CULT):** The concept of culture is complex and multi-dimensional, and contains many different aspects, for example, religion, social structure, language, political institutions, education and economics [20]. The behavioural intentions of Arab users are very much influenced by social values, interpersonal relationships and other issues related to religion [21]. CULT was measured by questions relating to

central cultural aspects including Image, Resistance to Change and Interpersonal social networks (*wasta* or social connection and nepotism).

With respect to the influence of Image, participants were asked if they felt that using m-Government services would enhance their social status and make them feel more sophisticated; and (82.2%) agreed that it would. Participants believed that using mobile government services would reduce the influence of interpersonal networks (*wasta*) (85.2%) and prevent any negative influence on their transaction by uncooperative employees (87.9%). Concerning Resistance to Change, only 27.2% of participants agreed that face-to-face dealings were better than using m-government, 17.4 % of them were neutral, and 55.4% disagreed entirely. Furthermore, 21.8% of participants agreed that visiting agencies to track transactions was preferable to online tracking, 64.4% disagreed and 13.8% were neutral. The composite of the CULT factor was 1.9997, a result that indicated that CULT is influential on users' intention to adopt and use of m-Government services. This corresponds with findings by other studies in the literature, that revealed that social and cultural aspects were significant influences on Saudis' intention to adopt and use e-Government systems from both citizens' and employee's perspectives [22]; and that the main barriers to Omanis adopting e-Government were cultural rather than technical issues [23].

**Perceived Trust (PT):** Different aspects of PT such as risks to privacy (sharing and storing personal information), security and trust were measured. Participants were asked if they felt that the Internet was not safe to be used for dealing with the government; and 41.0% believed that the Internet was safe, while 33.3 % believed it was safe, and 25.7% were neutral. This indicates some distrust in the Internet. By contrast, 73.9% of participants agreed that mobile government services were a safe and trustworthy environment in which to conduct government transactions, 20.8% were neutral and 5.3 % disagreed. This indicates that citizens' trust of m-Government services is high, and this might well have a positive influence on citizens' intention to use m-Government services. However, regarding whether providing personal information was safe 42.6% agreed, 19.8% were neutral and 37.6% disagreed; and 33% agreed their data could be misused when stored by m-Government systems, 24.3% were neutral and 42.8% disagreed. The total score for PT was 2.3586, indicating that PT is influential on intention to use m-Government in line with numerous studies that have noted that perceived trust was a significant factor [18], [24].

**Social Influence (SI):** This factor addresses users' perception of the effect of social influence and how this would encourage intention to adopt and use m-Government services. A large majority (80.5%) agreed that people important to them would think they should use mobile government services, 15.8% of them were neutral, and 3.7%

disagreed. 89.3% of participants said they would be encouraged to use m-Government services by their families and friends. The vast majority of participants (92.3%) intended to use m-Government services because it was the current trend. The composite of the SI factor was 1.6516, which indicates that SI is very influential on users' adoption and use; a result consistent with previous studies [24]. In a context like Saudi Arabia, where communities are very close, SI is a very important factor to include in the MGAUM, interestingly the desire to be seen by significant others as 'following the trend' was revealed as a powerful incentive in this context.

**Perceived Compatibility (PCOM):** This factor focuses on how users perceive the compatibility of m-Government services with their lifestyle and behaviour, and how this affects and encourages their intention to use m-Government services.

The vast majority (91.4%) of participants believed that using m-Government services would fit well with their lifestyles as well as being the way they liked to conduct government transactions (89.7%). The total score for PCOM was 1.5708 which indicates that this factor is very influential on intention to use; and that a high level of compatibility with the innovation would increase users' intention to adopt and use it. This finding is in line with other studies [24], [17]. Unlike previous research, this study included a high number of Saudi female participants. In a society where contact between sexes is sensitive for religious and cultural reasons, conducting government transactions on their phones arguably gives Saudi women both privacy and removes the need for face-to-face interactions with male government officials.

**Awareness (AW):** is the first stage where users experience a new service offered by the government. Participants were asked about which advertising methods could affect citizens' awareness of m-Government services and encourage their intention to adopt and use.

Many participants (76.7%) believed that they had a good level of knowledge about the benefits, features and services of m-Government, 14.9% were neutral, and 8.4% did not. Almost 68.6% of participants agreed it was easy to find out if a government agency offered its services via mobile devices, 17% of participants were neutral and 14.4% disagreed. Furthermore, 68.8% of participants agreed they had received enough information and guidance on how to use mobile government services, 14.9% disagreed and 16.3% were neutral. Also 66.6% were satisfied with the current awareness campaigns and advertising about m-Government services in Saudi Arabia, 16.4% were neutral and 17.0% were not satisfied. It is thus likely that AW positively influences citizens' intention to adopt and use m-Government services, in line with other findings [17], [25].

**Citizen Service Quality (CSQ):** Three main service quality dimensions were used to measure service quality in this study, i.e. reliability, responsiveness and empathy; and designed to measure customers' evaluation of the overall experience of services and explaining the difference between

users' perceptions and their expectations of the services offered by the government.

Participants were asked to what extent that they believed that m-Government service providers give 'a prompt service with a good response'; 75.6% agreed they did, 19.2% were neutral, and 5.2% disagreed. Similarly, most participants believed that m-Government service providers offered helpful assistance through SMS.

When asked whether they believed that information provided through m-Government services was accurate. 76.4% agreed it was, 19.9% were neutral and 3.7% disagreed. Furthermore, 69.4% of participants believed that m-Government service providers showed a sincere interest in solving citizens' problems, 24.8% were neutral, and 5.8% disagreed. The result echoes findings in Jordan [26] and Saudi Arabia [27] that the service quality dimension impacts significantly on citizen satisfaction and behavioural intention.

**System Quality (SQ):** This includes the technical aspects that are recognised by users, and which can affect their willingness and intention to adopt and use m-Government services. Regarding whether the speed of launching m-Government services applications or websites would affect participants' intention to use it, 79.1% agreed it would, 13.5% were neutral and 7.5% disagreed.

81.6% believed that m-Government was easy to navigate and that it provided good navigation functions; in contrast, 14.2% were neutral and only 4.2% disagreed. In respect of the existence of technical errors, such as applications crashing, links not working and unresponsiveness, (application/website), as well as the bad layout and unattractive interfaces of m-Government services 79.7% and 73.3% of participants respectively, agreed that these elements would reduce their willingness to use it. 86.2% of participants agreed that services should be compatible with mobile devices such as GPS and camera. Furthermore, 80.2% and 76.8% respectively believed that m-Government services provided fast responses to their enquiries as well as up-to-date information. The composite of the SQ factor was 1.8897, which indicated that SQ is influential on users' intention to adopt and use m-Government services. The results echo studies that indicated that when the system quality of m-Government services increases, citizens' intention to adopt and use them will also increase [26], [27].

**Perceived Mobility (PM):** The majority (89.8%) of participants expected to be able to use m-Government services anywhere and at any time; and found mobile government services were easily accessible, portable and easy-to-use on different models of Smartphone. 94.6% agreed it was important to get critical alert notifications on their mobiles from government agencies via text or email regarding passport renewal, traffic penalties and emergency cases whilst they were on the move. The composite of the PM factor was 1.5625 which indicates that PM is very influential on the intention to adopt and use m-Government services, and that citizens in Saudi Arabia value the ability to constantly access government services and information from

any location. The findings for this factor are consistent with previous studies in China [28].

The statistical analysis and discussion above allow us to estimate the most influential factors for increasing the intention to use m-Government services in Saudi Arabia.

#### IV. CONCLUSION

The researchers constructed the MGAUM model to identify factors revealed by the literature and personal professional experience to be likely to influence Saudi citizens' intention to adopt and use m-Government services.

The result of descriptive analysis presented in this paper indicates that all the proposed factors in MGAUM model (PU, PEOU, CULT, SI, PCOM, PT, AW, CSQ, SQ and PM) were statistically significant factors in influencing Saudi citizens' intention to adopt and use m-Government services and when properly addressed could increase the adoption rate of m-Government services.

We intend the results to provide a valuable insight into the main factors that influence citizen intention to adopt and use m-Government services in Saudi Arabia; which will be useful for researchers, the ICT industry and for policymakers who are keen to find strategies that result in quicker and more efficient take-up of such services.

This study added several contributions to theory and practice in the field of m-Government adoption and use. Firstly, this research developed the Mobile Government Adoption and Utilization Model (MGAUM) to analyse factors that affect users' adoption and use of m-government. MGAUM integrates the Technology Acceptance Model with a number of social, cultural and technological factors, taken from other recognized theoretical acceptance models that have been identified as key factors in the literature. Secondly, the MGAUM is empirically tested and validated by collecting and analysing primary data from the citizens' perspectives. Thirdly, this is one of the first few studies investigating the adoption and utilization of m-government in Saudi Arabia.

#### REFERENCES

- [1] S. Alghamdi, N. Beloff, Towards a comprehensive model for e-Government adoption and utilisation analysis: The case of Saudi Arabia, 2014 Fed. Conf. Comput. Sci. Inf. Syst. FedCSIS 2014. 2 (2014) 1217–1225. doi:10.15439/2014F146.
- [2] O. Al-Hujran, M.M. Al-Debei, A. Chatfield, M. Migdadi, The imperative of influencing citizen attitude toward e-government adoption and use, *Comput. Human Behav.* 53 (2015) 189–203. doi:10.1016/j.chb.2015.06.025.
- [3] H.J. Scholl, The mobility paradigm in electronic government theory and practice: A strategic framework, in: *Euro Mob. Gov. (Euro MGov) Conf.*, 2005: pp. 1–10.
- [4] S. Alotaibi, D. Roussinov, Developing and Validating an Instrument for Measuring Mobile Government Adoption in Saudi Arabia, *World Acad. Sci. Eng. Technol. Int. J. Soc. Behav. Educ. Econ. Bus. Ind. Eng.* 10 (2016) 710–716.
- [5] I. Kushchu, M.H. Kuscü, From E-government to M-government: Facing the Inevitable, *Proc. 3rd Eur. Conf. EGovernment.* (2003) 253–260. <http://citeseerx.ist.psu.edu>.
- [6] M. Ntaliani, C. Costopoulou, S. Karetso, Mobile government: A challenge for agriculture, *Gov. Inf. Q.* 25 (2008) 699–716. doi:10.1016/j.giq.2007.04.010.
- [7] I. Almarashdeh, M.K. Alsmadi, How to make them use it? Citizens acceptance of M-government, *Appl. Comput. Informatics.* 13 (2017) 1–6. doi:10.1016/j.aci.2017.04.001.
- [8] L.C. Serra, L.P. Carvalho, L.P. Ferreira, J.B.S. Vaz, A.P. Freire, Accessibility Evaluation of E-Government Mobile Applications in Brazil, *Procedia Comput. Sci.* 67 (2015) 348–357. doi:10.1016/j.procs.2015.09.279.
- [9] R. Lallana, E-government for development m-government: Mobile/wireless applications, in: 2004.
- [10] A.M. Alsenaidy, T. Ahmad, A Review of Current State M Government in Saudi, *Glob. Eng. Technol. Rev.* (2012) 5–8.
- [11] Y. Liu, H. Li, V. Kostakos, J. Goncalves, S. Hosio, F. Hu, An empirical investigation of mobile government adoption in rural China: A case study in Zhejiang province, *Gov. Inf. Q.* 31 (2014) 432–442. doi:10.1016/j.giq.2014.02.008.
- [12] M. Alonazi, N. Beloff, M. White, MGAUM — Towards a Mobile Government Adoption and Utilization Model: The Case of Saudi Arabia, *Int. J. Business, Hum. Soc. Sci.* 12 (2018) 459–466.
- [13] N. Blunch, *Introduction to Structural Equation Modelling Using IBM SPSS Statistics and AMOS*, Second Edition, SAGE publication Ltd, 2013.
- [14] Viswanathan, Madhubalan, S. Sudman, M. Johnson, Maximum versus meaningful discrimination in scale response: Implications for validity of measurement of consumer perceptions about products, *J. Bus. Res.* 75 (2004) 108–124. doi:10.1016/S0148-2963(01)00296-X.
- [15] V. Venkatesh, M.G. Morris, G.B. Davis, F.D. Davis, User Acceptance of Information Technology: Toward a Unified View, *MIS Q.* 27 (2003) 425–478. doi:10.1017/CBO9781107415324.004.
- [16] W.A. Alfarrar, *Analysing Questionnaires Data Using SPSS. Programs and Foreign Affairs Department, World Assembly of Muslim Youth.*, [Arabic Source]. (2009). <https://www.kantakji.com/media/9166/edu.pdf>.
- [17] H. Abdelghaffar, Y. Magdy, The Adoption of Mobile Government Services in Developing Countries: The Case of Egypt, *Int. J. Inf. Commun. Technol. Res.* 2 (2012) 333–341.
- [18] S.Y. Hung, C.M. Chang, S.R. Kuo, User acceptance of mobile e-government services: An empirical study, *Gov. Inf. Q.* 30 (2013) 33–44. doi:10.1016/j.giq.2012.07.008.
- [19] S. Alghamdi, N. Beloff, Exploring Determinants of Adoption and Higher Utilisation for E-Government: A Study from Business Sector Perspective in Saudi Arabia, 5 (2015) 1469–1479. doi:10.15439/2015F257.
- [20] L. Chang, Cross-Cultural Differences in International Management Using Kluckhohn-Strodtbeck Framework, *J. Am. Acad. Bus.* 2 (2002) 20–27.
- [21] P.J.-H. Hu, S.S. Al-Gahtani, H.-F. Hu, Arabian workers' acceptance of computer technology: A model comparison perspective, *J. Glob. Inf. Manag.* (2014). doi:10.4018/jgim.2014040101.
- [22] R. HMBP, Role of National Culture on the Use of e-Government Services in Sri Lanka, *J. Bus. Financ. Aff.* 5 (2016). doi:10.4172/2167-0234.1000182.
- [23] S.J. Naqvi, H. Al-shihi, M-Government Services Initiatives in Oman, 6 (2009).
- [24] C. Sellitto, M.W.L. Fong, An investigation of mobile payment (m-payment) services in Thailand, *Asia-Pacific J. Bus. Adm.* 8 (2015) 37–54. doi:10.1108/APJBA-10-2014-0119.
- [25] S.A. Al-Somali, R. Gholami, B. Clegg, An investigation into the acceptance of online banking in Saudi Arabia, *Technovation.* 29 (2009) 130–141. doi:10.1016/j.technovation.2008.07.004.
- [26] F.T. Qutaishat, Users' Perceptions towards Website Quality and Its Effect on Intention to Use E-government Services in Jordan, *Int. Bus. Res.* 6 (2012) 97–105. doi:10.5539/ibr.v6n1p97.
- [27] A.M. Baabdullah, A.A. Alalwan, N.P. Rana, H. Kizgin, P. Patil, Consumer use of mobile banking (M-Banking) in Saudi Arabia: Towards an integrated model, *Int. J. Inf. Manage.* 44 (2019) 38–52. doi:10.1016/j.ijinfomgt.2018.09.002.
- [28] Y.S. Yen, F.S. Wu, Predicting the adoption of mobile financial services: The impacts of perceived mobility and personal habit, *Comput. Human Behav.* 65 (2016) 31–42. doi:10.1016/j.chb.2016.08.017.



# Developing a Model and Validating an Instrument for Measuring the Adoption and Utilisation of Mobile Government Services Adoption in Saudi Arabia

Mohammed Alonazi  
University of Sussex  
Informatics Department  
Brighton, UK  
Email: M.Alonazi@sussex.ac.uk

Natalia Beloff  
University of Sussex  
Informatics Department  
Brighton, UK  
Email: N.Beloff@sussex.ac.uk

Martin White  
University of Sussex  
Informatics Department  
Brighton, UK  
Email: M.White@sussex.ac.uk

**Abstract**—Many governments worldwide are taking advantage of the latest developments in mobile technology to take the digital delivery of government information and services (e-government) to their citizens a stage further. Accessing government information and services via a mobile device is known as m-government, a system designed to serve citizens, companies and government agencies alike. M-government also has unique advantages over e-government, not least enabling users to access government services at any time and from any location. This paper presents a pilot study of the MGAUM model that was developed to analyze factors influencing the adoption rate of m-government services in Saudi Arabia. With the aim of validating a survey instrument with which to conduct the main study in Saudi Arabia, a pilot survey instrument was developed and modified by using previous instruments from research into both e-government and m-government. This pilot questionnaire was distributed to 71 Saudi citizens in the UK, and the final sample was 58 valid responses. The results of this pilot study reveal that all items in the survey instrument were reliable and valid within the range of acceptable academic research and suitable for use in the main study. The results of this pilot study were analyzed using SPSS.

## I. INTRODUCTION

The revolution in Information and Communication Technologies (ICT) has resulted in governments across world digitizing their services to citizens through tools such as mobile technologies and the Internet. Arguably, the implementation of e-government has resulted in better communication between citizens and their governments with enhanced access to services and information as well as improvements in government transparency, accountability and public governance[1], [2], [3]. Several governments have also implemented mobile government (m-government) by developing innovative service delivery channels that utilize wireless and mobile technologies [4], [5]. Accordingly, in this study, m-Government is defined as the use of mobile technology to deliver and improve e-government services and information to citizens, businesses and all government agencies. Previous studies have identified m-government as both separate to e-government or as an extension or replacement of it [5], [6]. Kushchu & Kuscu stress that it is an advance in government service delivery as it offers the public a valuable extra means to access services and information[4]. Although the same principles are common to

IEEE Catalog Number: CFP1985N-ART ©2019, PTI 633

e-government and m-government, the latter is distinguished by features that are particular to it:

- The main advantage for citizens of m-government is its mobility, giving citizens access to the network at any time and from anywhere[7].
- As mobile phones can be used easily, citizens can instantly receive messages from government service providers[8].
- In many countries, the mobile phone has recently become the primary way people communicate over distance; the mobile phone has thus arguably become a part of everyday life for many people[9], [10]. For this reason, being able to access government services via mobile devices might be the best route for citizens.
- More and more people are using their mobiles to access the web. For citizens, this means they neither have to visit the service provider in person nor go home to use their computer when they want to access government services and information[8], [10].
- Access to the Internet may depend on a country's economy in terms of the extent of Internet access and how many citizens have access to computers[6],[4]; these limitations can be overcome by providing mobile services.

Therefore, by using an m-government system provided by wireless technology, citizens will acquire opportunities for personalized, real-time access to government services and information[11], [7]. Users in remote areas will especially benefit from being able to access government services and information as m-services have the advantage of being affordable, easy and immediate to access and require a relatively low level of digital literacy to operate successfully [12], [7]. Given these characteristics, the adoption of an m-government system has benefits for citizens and government alike. Any government services that citizen can use via mobile devices, such as apps, mobile website (web platform), SMS or call centre, are considered as mobile government services. So, this research focuses on all government services are provided via mobile devices. Therefore, the government agencies should concentrate on

offering different channels for citizens to conduct government transactions at the same time to allow fast and easy access to these services fast. However, the main difference between the website and mobile devices is that features on the mobile web platform and Apps that add value to services, such as reminders, location and camera, are not found on the website. However, from the researcher's experience, people prefer to use native applications rather than a mobile responsive website.

The goals of this research are to investigate and understand Saudi citizens' perceptions towards the adoption and utilization of m-government services in developing countries, particularly Saudi Arabia, in order to increase the adoption rate of m-government services. This paper focused on validating the MGAUM model that has been developed by integrating the TAM model in conjunction with social, cultural and technological factors that the available literature has suggested are key to understanding technology acceptance within a specific context. The results of this pilot study reveal that all items in the survey instrument were reliable and valid within the range of acceptable academic research and suitable for use in the main study.

## II. BACKGROUND AND CONTEXT

Despite m-government systems being available for several years, citizens' adoption of e-government services in general and m-government services in particular still falls below expectations [1],[2],[13],[14]. Furthermore, in Saudi Arabia, like in most developing countries, m-government implementation is still in its infancy and there are many challenges related to implementation, adoption and use [1],[15],[16]. Factors such as the rate of mobile device and Internet penetration and their security, reliability and effectiveness, affect how successful a government will be at implementing m-government and user adoption and accounts for global variation [11]. However, there is a lack of research that allows a clear understanding of how factors such as these might impact the adoption and use of m-government services. This study rectifies this problem by providing a theoretical model purposely developed to carry out empirical research in this area. The results of this research will yield new insights about the key factors influencing the adoption of Saudi m-government, which will be invaluable to policy makers who require strategies that will result in faster and more efficient adoption of m-government services; as well as providing new information for researchers in the field and the ICT industry.

Research carried out in a number of different areas such as Malaysia and rural China [12], [17],[18] have made use of adaptations of the Technology Acceptance Model (TAM) and provided examples of how a number of social, cultural and technical factors can usefully be included in the TAM to

provide insights into the influences on citizens' intention to utilize m-government systems to access services and information. Cultural and technological factors like culture, trust and lack of necessary infrastructure have been demonstrated to be significant by comparative studies of m-government adoption in developing and developed countries [19],[20].

The adoption of m-government in Arab countries, however, still requires further research. Studies conducted in these areas [21],[22],[23],[20],[24] have revealed that factors such as trust, citizens' perceptions of the compatibility of m-government with their lifestyles, culture, awareness and the quality of the system are significant. Further, these studies show that there have been no empirical studies of m-government adoption in Saudi Arabia that includes factors like compatibility or culture. Similarly, there are no studies that take the quality of both technical and human factors into account or that investigate the issues from the viewpoint of the providers in addition to the intended users. There is clearly a need to carry out further research into Saudi m-government adoption.

In order to analyze factors that affect users' adoption and use of m-government, the researcher has developed a model called the Mobile Government Adoption and Utilization Model (MGAUM)[25]. The remainder of this paper consists of the following sections: Section 3 outlines the MGAUM, Section 4 covers the research methodology; Section 5 contains a descriptive and statistical analysis of the findings and Section 6 is the conclusion.

## III. THE PROPOSED RESEARCH MODEL (MGAUM)

In order to analyze factors that affect users' adoption and use of m-government, the researcher has developed a model called the Mobile Government Adoption and Utilization Model (MGAUM)[25]. MGAUM has been developed based on a critical analysis of the literature that relates to acceptance of technology, in conjunction with insights from several models and theories that are commonly used to analyze acceptance and usage of technologies. MGAUM integrates the Technology Acceptance Model [26], with a number of social, cultural and technological factors, taken from other recognized theoretical acceptance models that have been identified as key factors in the literature.

Further, MGAUM contains one dependent variable namely: Intention to use m-government (ITU), and three groups of independent variables namely: Practical Factors (PF), Human Factors (HF), and Technical Factors (TF). These independent variables comprise the key factors that critically influence the adoption and use of m-government.

This model uses the TAM as a starting point along with factors from other theories to analyze how users adopt m-

government services. Two factors were taken from the TAM: Perceived Ease of Use (PEOU) and Perceived Usefulness (PU). These two factors comprise the Practical Factors of MGAUM; the Human Factors are those that were revealed by the literature as important in further understanding the role played by the individual users' acceptance in this context and the Technical Factors are those identified as most likely to affect the adoption of m-government. The final design of MGAUM also used the researcher's experience of the local problems of accessing government services and information in Saudi Arabia to select those factors identified as key in the relevant literature. The main aim of MGAUM is to investigate the adoption of m-government services by citizens in developing countries, particularly Saudi Arabia, in order to increase the adoption rate of m-government services. The research model MGAUM is shown in Fig. 1.

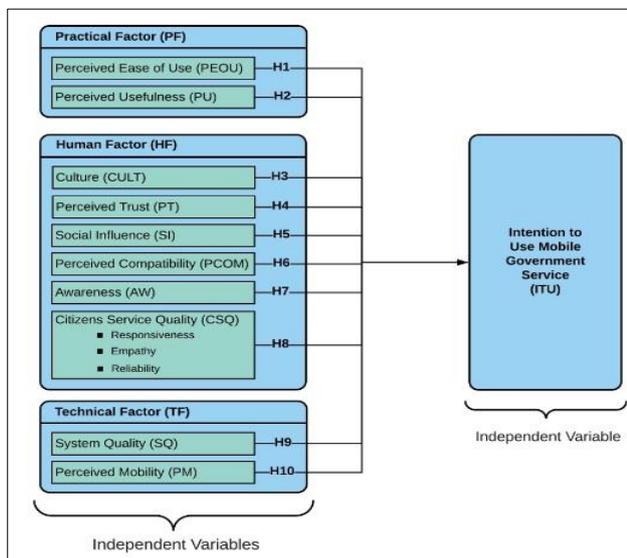


Fig. 1: The research model (MGAUM) [25].

The hypotheses of the study are as follows[25]:

H1: Perceived Ease of Use positively influences intention to use m-government services.

H2: Perceived Usefulness positively influences intention to use m-government services.

H3: Culture influences intention to use m-government services.

H4: Trust positively influences intention to use m-government services.

H5: Social Influence affects intention to use m-government services.

H6: Compatibility positively influences intention to use m-government services.

H7: Awareness positively influences intention to use m-government services.

H8: Citizen service quality factors (responsiveness, empathy and reliability) positively influence intention to use m-government services.

H9: System Quality positively influences intention to use m-government services.

H10: Perceived Mobility positively influences intention to use m-government services.

#### IV. THE RESEARCH METHODOLOGY

This pilot study was conducted in the United Kingdom. The questionnaire was distributed to Saudi citizens (public users) whether they had used mobile government services or not. Participants had to be 18 years or older to participate in this survey. Seventy-one participants took part in the questionnaire; but as all the questions had to be answered as they represented the research model, incomplete questionnaires were excluded from the survey. Therefore, the total valid responses constituted a sample of 58 Saudi citizens. The pilot study data were analysed by using the SPSS program.

The survey questionnaire contains 76 items and all the questions were developed and modified from instruments used in previous research into both E-government and M-government. All items were measured with a 5-point Likert scale from 'strongly agree' to 'strongly disagree'. In order to collect enough data to ensure a thorough analysis, the questionnaire had to be relatively lengthy. Manual distribution meant that the researcher could clarify points for the participants if necessary and contributed to obtaining a higher response rate.

##### A. Reliability and validity of the study

If a research instrument is to be reliable, it should produce similar results if used in comparable conditions [27]. Furthermore, an instrument's reliability depends on how free it is of error [28]. In order to assess the reliability of the questionnaire used in the pilot study, Cronbach's Alpha was chosen because the internal consistency of the constructs used in the questionnaire had to be established; in other words, we needed to measure the extent to which items in the questionnaire measured the same things when referring to a specific independent or dependent construct, and how these related to each other. Cronbach's Alpha is the most commonly used test to calculate and evaluate internal consistency, and thus reliability [29]. Cronbach's Alpha has a scale of 0 to 1, with 1 being the highest reliability, a value of independent and dependent variables 0.6 is considered to be acceptable [30]. Table 1 indicates the Cronbach's Alpha results for the complete pilot study instrument, which demonstrates that the reliability of each of the constructs (independent and dependent variables) lies within the range of what is considered acceptable in academic research.

A valid research instrument measures what the researcher intended [31], and the validity of an instrument is the extent to which it does this and provides the information required [32]. In order to establish the validity of our research instrument, the face validity and content validity methods

were selected rather than construct validity and criterion validity. This method is designed to establish the extent to which the purpose of the instrument is clear even to the lay person with only basic education [33], for example a 1st grader at school. There is a high level of face validity if the items in the questionnaire are clear and unambiguous, if the items are perceived as difficult to understand or confusing, then the face validity is low [33]. The pilot study allowed the face validity of each item to be tested by participants so as to ensure that the model's factors measured what they were intended to measure. Items without sufficient clarity, unambiguity or relevance were revised or deleted accordingly. Furthermore, six academics, all of whom had expertise in the field, were asked to review the items in the research instrument, this review and the pilot study demonstrated that all the items used therefore had a good degree of content validity.

TABLE 1: INTERNAL CONSISTENCY OF THE PILOT STUDY INSTRUMENT

Measured Variable	No. of Items	Cronbach's Alpha
Perceived Usefulness (PU)	7	.841
Perceived Ease of Use (PEOU)	4	.848
Social Influence(SI)	3	.753
Perceived Compatibility (PCOM)	2	.824
Perceived Trust (PT)	7	.618
Culture (CULT)	5	.606
Awareness (AW)	4	.816
Perceived Mobility (AW)	3	.819
Citizens Service Quality (CSQ)	8	.920
System Quality (SQ)	7	.755
Intention to use (ITU)	4	.879

## V. DESCRIPTIVE AND STATISTICAL ANALYSIS

### A. Respondents' demographic data

In this study, 58 participants answered the questionnaire, and that 38 participants were male (65.5 %) and 20 were female (34.5 %). The highest percentage of participants was in the 31–45 age group; the largest number held a Bachelor degree and over half the participants (51.7%) were government employees. All participants had smartphone devices, with a large majority of participants using mobiles and the Internet in daily life (93.1% and 94.8% respectively). Approximately three-quarter of the respondents (74.1 %) have some knowledge about m-government services in Saudi Arabia, whereas 25.9% had no knowledge. Moreover, the majority of participants (86.2%) already used M-government services, but 13.8% had never used it. Also, the survey asked the participants that already used m-government services to rate their general experience. The result showed that 44.4% were very satisfied with m-government services, 52% were satisfied to some extent, with only 4% not satisfied with m-government services. Furthermore, 27.6% of participants

reported that the requirements of the intended m-government services were not clear and 34.5% of them reported that the system quality of m-government services was not good.

When asked about what advertising methods could affect awareness of m-government services and encourage their use, participants rated them as follows: social media, emails and text messages (20.17%), advertisements in public areas (17.54%), TV and Radio channels (17.1 %), government agencies' websites (14.03 %) and finally, newspapers and magazines (14.03 %).

### B. The Correlation Analysis

A correlation coefficient analysis was run on this pilot study to discover the relationships between all constructs and to establish their significance. In 1988, Cohen proposed a guideline for correlation coefficient values as follows: Strong  $r = .50$  to  $1.0$ , Moderate  $r = .30$  to  $.49$  and Weak  $r = .10$  to  $.29$  [34]. The correlation coefficients can be seen in Table 2, which shows there is a positive relation between the intention to use m-government (Dependent variable ITU) construct and the rest of the constructs, with differences in the strength of this relationship from one construct to another. For instance, there is a strong correlation between Perceived Usefulness and Perceived Compatibility with Intention to use m-government services, with moderate correlation for Perceived Ease of Use, Social Influence and Culture whereas there was only a low correlation for Perceived Trust, Awareness, Perceived Mobility, Citizens Service Quality and System Quality.

TABLE 2 :THE CORRELATION BETWEEN THE VARIABLES

		Correlations										
		INT	PU	PEOU	SI	PCOM	PT	CULT	AW	PM	CSQ	SQ
INT	Pearson Correlation	1	.682**	.479**	.411**	.598**	.247	.323*	.089	.251	.074	.279*
	Sig. (2-tailed)		.000	.000	.001	.000	.062	.013	.506	.057	.581	.034
	N	58	58	58	58	58	58	58	58	58	58	58
**. Correlation is significant at the 0.01 level (2-tailed).												
*. Correlation is significant at the 0.05 level (2-tailed).												

The correlation results show that user experience is the first step in adoption; and if a user finds m-Government services easy to use and that it saves time, effort and is compatible with lifestyle and mobile devices, then this impacts positively on his/her behavioural intention to adopt and use them. If services are easy to use, and people do not have to rely on asking for help from another person to use m-government service, the number of users will increase. Thus, PU, PEOU and PCOM are essential factors in this study.

The result also indicated that government agencies should focus on the following factors (PT, CULT, AW, PM, CSQ, SI and SQ) to motivate and increase citizens' intention to adopt and use m-government services. Therefore, the government needs to raise awareness about the main goals of m-Government, the availability m-Government services and

the advantages and benefits gained from the use of m-Government services to conduct various transactions. Public awareness could be enhanced in various ways including interactive advertising and social media campaigns as well as traditional advertising methods such as brochures and advertisements on TV, public transport and in newspapers.

The pilot study has demonstrated that the initial design of the questionnaire will be suitable for the main study to be carried out in situ in Saudi Arabia. The result of the main study will be compared to the pilot study result for extra validation.

## VI. CONCLUSION

This pilot study focused on validating the MGAUM model that has been developed by integrating the TAM model in conjunction with social, cultural and technological factors that the available literature has suggested are key to understanding technology acceptance within a specific context. We intended in this study to provide a valuable insight into the main factors that influence adoption of m-government services in Saudi Arabia, which will be useful for researchers, the ICT industry and for policymakers who are keen to find strategies that result in quicker and more efficient take-up of such services. The reliability and validity of the pilot study have been established. The findings of this pilot study provided us with the basic information and will open the doors for future discussions. We plan to conduct this study in Saudi Arabia with a large number of Saudi citizens to explore what factors encourage them to use m-government services and what the barriers to acceptance are.

## VII. REFERENCES

- [1] S. Alghamdi, N. Beloff, Towards a comprehensive model for e-Government adoption and utilisation analysis: The case of Saudi Arabia, 2014 Fed. Conf. Comput. Sci. Inf. Syst. FedCSIS 2014. 2 (2014) 1217–1225. doi:10.15439/2014F146.
- [2] O. Al-Hujran, M.M. Al-Debei, A. Chatfield, M. Migdadi, The imperative of influencing citizen attitude toward e-government adoption and use, *Comput. Human Behav.* 53 (2015) 189–203. doi:10.1016/j.chb.2015.06.025.
- [3] S.Y. Hung, C.M. Chang, S.R. Kuo, User acceptance of mobile e-government services: An empirical study, *Gov. Inf. Q.* 30 (2013) 33–44. doi:10.1016/j.giq.2012.07.008.
- [4] I. Kushchu, M.H. Kusecu, From E-government to M-government: Facing the Inevitable, *Proc. 3rd Eur. Conf. EGovernment.* (2003) 253–260. <http://citeseerx.ist.psu.edu>.
- [5] H.J. Scholl, The mobility paradigm in electronic government theory and practice: A strategic framework, in: *Euro Mob. Gov. (Euro MGov) Conf.*, 2005: pp. 1–10.
- [6] S. Alotaibi, D. Roussinov, Developing and Validating an Instrument for Measuring Mobile Government Adoption in Saudi Arabia, *World Acad. Sci. Eng. Technol. Int. J. Soc. Behav. Educ. Econ. Bus. Ind. Eng.* 10 (2016) 710–716.
- [7] M. Ntaliani, C. Costopoulou, S. Karetos, Mobile government: A challenge for agriculture, *Gov. Inf. Q.* 25 (2008) 699–716. doi:10.1016/j.giq.2007.04.010.
- [8] I. Almarashdeh, M.K. Alsmadi, How to make them use it? Citizens acceptance of M-government, *Appl. Comput. Informatics.* 13 (2017) 1–6. doi:10.1016/j.aci.2017.04.001.
- [9] L.C. Serra, L.P. Carvalho, L.P. Ferreira, J.B.S. Vaz, A.P. Freire, Accessibility Evaluation of E-Government Mobile Applications in Brazil, *Procedia Comput. Sci.* 67 (2015) 348–357. doi:10.1016/j.procs.2015.09.279.
- [10] R. Lallana, E-government for development m-government: Mobile/wireless applications, in: 2004.
- [11] A.M. Alsenaidy, T. Ahmad, A Review of Current State M Government in Saudi, *Glob. Eng. Technol. Rev.* (2012) 5–8.
- [12] Y. Liu, H. Li, V. Kostakos, J. Goncalves, S. Hosio, F. Hu, An empirical investigation of mobile government adoption in rural China: A case study in Zhejiang province, *Gov. Inf. Q.* 31 (2014) 432–442. doi:10.1016/j.giq.2014.02.008.
- [13] R. Alotaibi, L. Houghton, K. Sandhu, Exploring the Potential Factors Influencing the Adoption of M-Government Services in Saudi Arabia: A Qualitative Analysis, 11 (2016) 56–72. doi:10.5539/ijbm.v11n8p56.
- [14] N.P. Rana, Y.K. Dwivedi, Citizen's adoption of an e-government system: Validating extended social cognitive theory (SCT), *Gov. Inf. Q.* 32 (2015) 172–181. doi:10.1016/j.giq.2015.02.002.
- [15] S. Alotaibi, D. Roussinov, A conceptual model for examining mobile government adoption in Saudi Arabia, *Proc. Eur. Conf. e-Government, ECEG.* 2015-Janua (2015) 369–375.
- [16] K. Assar, M-government in Saudi Arabia, 5 (2015) 76–83.
- [17] T.M. Faziharudean, T. Li-Ly, Consumers' behavioral intentions to use mobile data services in Malaysia, *African J. Bus. Manag.* 5 (2011) 1811–1821. doi:10.5897/AJBM10.794.
- [18] M.A. Alqahtani, R.S. AlRoobaea, P.J. Mayhew, Building a Conceptual Framework for Mobile Transaction in Saudi Arabia: A User 's Perspective Building a Conceptual Framework for Mobile Transaction in Saudi Arabia: A User 's Perspective, (2014) 967–973. doi:10.1109/SAI.2014.6918303.
- [19] A. Al-Hadidi, Y. Rezgui, Adoption and Diffusion of m-Government: Challenges and Future Directions for Research, in: *IFIP Int. Fed. Inf. Process.*, 2010: pp. 88–94.
- [20] M.A. Shareef, Y.K. Dwivedi, S. Laumer, N. Archer, Citizens' Adoption Behavior of Mobile Government (mGov): A Cross-Cultural Study, *Inf. Syst. Manag.* 33 (2016) 268–283. doi:10.1080/10580530.2016.1188573.
- [21] N.A.S. Almuraqab, M-government adoption factors in the UAE: A partial least-squares approach, *Int. J. Bus. Inf.* 11 (2017) 404–431. <http://ijbi.org/ijbi/article/view/191>.
- [22] N.A.S. Almuraqab, S.M. Jasimuddin, Factors that Influence End-Users ' Adoption of Smart Government Services in the UAE: A Conceptual Framework, 20 (2017) 11–23.
- [23] M.A. Shareef, V. Kumar, U. Kumar, Y.K. Dwivedi, e-Government Adoption Model (GAM): Differing service maturity levels, *Gov. Inf. Q.* 28 (2011) 17–35. doi:DOI: 10.1016/j.giq.2010.05.006.
- [24] F.D. Davis, R.P. Bagozzi, P.R. Warshaw, User Acceptance of Computer Technology: a Comparison of Two Theoretical Models., *Manag. Sci.* Aug1989. 35 (1989) 982–1003. doi:10.2307/2632151.
- [25] M. Alonazi, N. Beloff, M. White, MGAUM — Towards a Mobile Government Adoption and Utilization Model: The Case of Saudi Arabia, *Int. J. Business, Hum. Soc. Sci.* 12 (2018) 459–466.
- [26] F.D. Davis, R.P. Bagozzi, P.R. Warshaw, User Acceptance of Computer Technology: a Comparison of Two Theoretical Models., *Manag. Sci.* Aug1989. 35 (1989) 982–1003. doi:10.2307/2632151.
- [27] N. Blunch, *Introduction to Structural Equation Modelling Using IBM SPSS Statistics and AMOS*", Second Edition, SAGE publication Ltd, 2013.
- [28] U. Sekaran, *Research methods for business*, fourth, John Wiley & Sons, Inc, 2003. doi:10.1017/CBO9781107415324.004.
- [29] E. Eucharria, O. Nnadi, *Health Research Design and Methodology*, Library of Congress, CRC Press, 1999.
- [30] D. Suhr, M. Shay, *Guidelines for Reliability , Confirmatory and Exploratory Factor Analysis*, in: University of Northern Colorado, Cherry Creek Schools, USA, 2008., 2008: pp. 1–15.
- [31] D.K. BHATTACHARYYA, *Cross-cultural Management: Texts and Cases.*, PHI learning private limited, New Delhi, 2010.
- [32] D. Colton, R.W. Covert, *Designing and Constructing Instruments for Social Research and Evaluation*, Jossey-Bass, San Francisco (an imprint of Wiley), 2007.
- [33] B. Nevo, FACE VALIDITY REVISITED, *J. Educ. Meas.* 22 (1985) 287–293. doi:10.1111/j.1745-3984.1985.tb01065.x.
- [34] Cohen J., *Statistical Power Analysis for the Behavioural Science*, 2nd Editio, Routledge, 1988.



## Towards Data Quality Runtime Verification

Janis Bicevskis  
Faculty of Computing  
University of Latvia, Latvia  
Email:  
Janis.Bicevskis@lu.lv  
ORCID: 0000-0001-  
5298-9859

Zane Bicevska  
DIVI Grupa Ltd,  
Latvia  
Email:  
Zane.Bicevska@di.lv  
ORCID: 0000-0002-  
5252-7336

Anastasija Nikiforova  
Faculty of Computing  
University of Latvia, Latvia  
Email:  
Anastasija.Nikiforova@lu.lv  
ORCID: 0000-0002-  
0532-3488

Ivo Oditis  
DIVI Grupa Ltd,  
Latvia  
Email:  
Ivo.Oditis@di.lv  
ORCID: 0000-0003-  
2354-3780

□ **Abstract**— This paper discusses data quality checking during business process execution by using runtime verification. While runtime verification verifies the correctness of business process execution, data quality checks assure that particular process did not negatively impact the stored data. Both, runtime verification and data quality checks run in parallel with the base processes affecting them insignificantly. The proposed idea allows verifying (a) if the process was ended correctly as well as (b) whether the results of the correct process did not negatively impact the stored data in result of its modification caused by the specific process. The desired result will be achieved by use of domain specific languages that would describe runtime verification and data quality checks at every stage of business process execution.

**Keywords**—data quality, runtime verification, business process, domain specific languages

### I. INTRODUCTION

NOWADAYS, the most part of processes is based on more than one information system or service. Moreover, the environment where processes are running usually is very heterogeneous. As a result, besides users, other information systems and changes made in them may affect execution of the initial process. One of the typical solutions for such situations is detection of incorrect execution by system monitoring or support staff, however, identification of affected business processes isn't possible in this case. As a result, the necessity for runtime verification of business processes appears to keep the process consistent at any time [1]. As it was discussed in [2], runtime verification of business processes allows (a) detection of incorrect execution that is possible in the case of system monitoring, but also (b) identification of business processes that may be affected by it. As a result, asynchronous runtime verification of business process was proposed focusing on timely and accurate problem identification.

□ The research leading to these results has received funding from the research project "Competence Centre of Information and Communication Technologies" of EU Structural funds, contract No. 1.2.1.1/18/A/003 signed between IT Competence Centre and Central Finance and Contracting Agency, Research No. 1.7 "The use of business process models for full functional testing of information systems".

However, runtime verification of business processes checks only the correctness of process execution based on the evaluation of process execution sequence, meanwhile this research proposes to extend it by involving data quality mechanism, that will assure the specific process did not negatively affect data stored in information systems that were affected by this process. It is achieved by applying data object-driven approach to data quality evaluation [3]. This approach is based on definition of data object which quality should be analysed, quality requirements definition that are applied to the parameters of the defined data object and measuring data quality. In scope of the proposed solution, data object is derived from data that were affected by running process. Data quality requirements are defined to check whether data are still correct and "external constraints" are still valid.

The paper deals with following issues: concepts used for data quality checks in the runtime verification (Section 2), idea and main concepts of the proposed solution (Section 3), analysis of the proposed solution (Section 4), conclusions and future work (Section 5).

### II. BASIC CONCEPTS

This chapter briefly discusses the concepts used in runtime verification and data quality research that are necessary for discussing the ideas and solutions proposed.

#### A. Runtime Verification

Runtime verification mechanism proposed in [2], doesn't intervene into execution of processes. It observes processes from the aside, collecting and verifying events confirming process step execution, in accordance with business process description. The main point is checking of the verification of business process execution in compliance with the process verification description. The description of the verification process must specify two aspects: (a) event confirming step completion, and (b) the time when each step in the process must be finished. Two main components of this mechanism are agents and controller.

The agent plays a role of event detector. It is software that checks the occurrence of a specific event. An example of such event can be record insertion. All events detected by

event agents are sent to the centralized controller for verification.

The controller analyses process verification descriptions, collecting event messages that were sent by agents and verifies flow compliance with the verification description.

Agents are developed for different components (databases, file systems, email servers etc.) and not implemented into software under verification. Thereby proposed mechanism allows verifying business processes executed by more than one system, running over several platforms, and even provided by more than one operator.

Runtime verification of business processes allows detection of incorrect execution and identification of business processes that may be affected by it. Both, detection of incorrect execution and identification of business processes that may be affected by it, take place immediately after changes were made by including appropriate checks in the runtime verification of business processes. In comparison with more traditional for such cases system monitoring, it allows to fix the occurrence of such problem as soon as possible for its timely solving to achieve as high result as possible. Moreover, identification of business processes that may be affected by it usually isn't considered at all. In other words, asynchronous runtime verification of business process focuses on timely and accurate problem identification.

Significant benefit of this approach is that it can be used when existing software does not respect any component addition. It is very useful in cases when the source code of some software is not available or there is not enough knowledge on all details of software implementation.

The idea of the proposed runtime verification is close enough to the formal class of runtime verification discussed in [5].

### B. Data Object-Driven Data Quality Model

Data object-driven data quality model consists of 3 main components: (1) data object that defines the data which quality must be analysed, (2) data quality specification that defines conditions which must be met to admit data as qualitative, and (3) quality evaluation process that defines the procedure that must be performed to evaluate data quality [6].

Every component of the quality model is represented by flowchart-based diagrams that are easy to read, create and edit. This approach is based on three domain specific languages (DSLs) created for every model's component.

As it follows from the listed components, used solution doesn't use the concept of "dimension". Instead, the wider concept of "data quality specification" is used. The main idea of this model is that all components are fully defined by user in correspondence with users' viewpoint on the specific dataset and quality.

Data object is defined in accordance with data needed to be analysed, the parameters that do not make sense for

particular users and use-cases are ignored. Data objects of the same structure form data object class where each individual data object may contain parameter values fully or partially [4]. Similarly, data quality specification is defined by user depending on the use-case. The nature of quality requirement or condition depends on the users' need. It can be compared with rule-based approaches used for relational database analysis. However, this approach reduces this limitation and can be applied to wider range of data structures. Currently, it can be applied to structured and semi-structured data. Data quality specification can be defined informally or in formal way, however at the last step all requirements are replaced by executable artefacts such as SQL statements or program code that further are executed.

Such approach is quite simple as it is very intuitive and close enough to "data" and "data quality" concepts nature. As a result, it is expected to be well-understood even by non-IT and non-data quality experts. It is one of the main benefits of this approach as usually approaches for data quality evaluation are suitable mostly for IT- and DQ- experts requiring deep knowledges in both areas [6]. However, data object-driven model can be used by wide audience without need to make in-depth analysis of the basics of the approach as it usually happens with other approaches where at least an exploration of the list of dimensions, their meanings and criteria under each of them, needs a lot of time for every particular solution as criteria differentiate from case to case [3], [6], [7]. At the same time, the proposed approach already demonstrated its effectiveness by applying it to real datasets [4], [6] – [8].

As a result, the given research uses data object-driven data quality model as the most appropriate option. In order to explain basics of the proposed solution, the next chapter summarizes the basic concepts involved in it. The required modifications are outlined to achieve desirable result.

## III. A PROPOSED SOLUTION

The main idea of the proposed solution is to allow check the quality of data while business process is executed after each data object update. In other words, when the process  $X_a$  step  $S_n$  is done, check data quality requirement  $DQS_1(X_a, S_n)$ . Data quality requirements in scope of this research are "external constraints" [9] defined for the particular process.

### A. Concept of Runtime Data Quality Control

Following runtime verification mechanisms proposed by [2], a business process description should be defined for process verification. Process description contains process states and events linking states (Fig. 1). From the data quality perspective each of processes may affect one or more data objects. Accordingly, data object class definition should be added to the process definition (described in the previous section). These objects are to be changed during business process execution and there could be different verification rules for each of process steps.

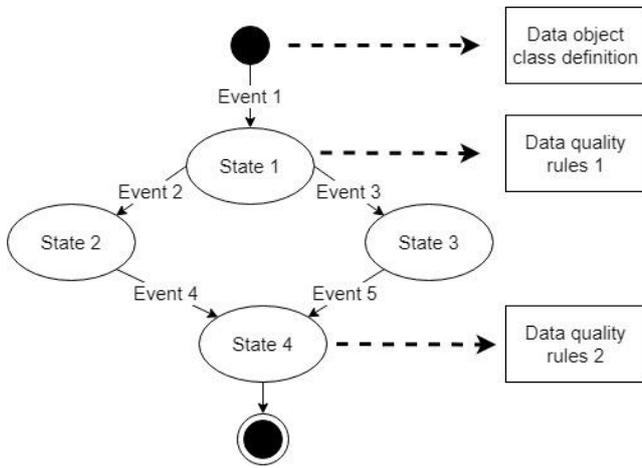


Fig. 1 Business process verification procedure

When process verification is running, new process verification instance is created by each business process start event. If necessary, corresponding data object is extracted verified according data quality definition. When next process execution event is detected, process verification instance is moved to the next state and next data object version is extracted, and its quality is verified. Thereby each process verification instance may have more than one data object instance and data quality of one data object may be verified not just at a fixed moment of time, but between its modifications accordingly (Fig. 1). This allows to identify (a) data quality loss exactly when it happens and (b) the incorrectly working process events.

According to [2] the business process verification involves two components: a *controller* and *agents*. Data quality verification adds *Data link* component to the solution (Fig. 2). Data link provides required connection to the database with business process objects and extracts data object copy when it is required by runtime verification controller. Therefore, not only data object class definition is required for data quality runtime verification, but also a definition of data object mapping to business process database: data link uses this definition for data object extraction from business process database.

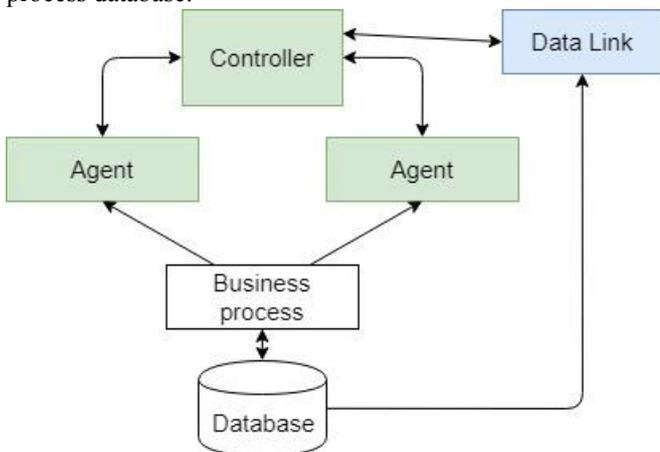


Fig. 2 Architecture of the proposed solution

**B. Data Object Definition**

DSL for defining of data objects is discussed in detail in [4] and [6]. As more and more customers are using electric scooter renting services, this will serve as an example to explain the proposed solution. One scooter rental case will be a data object sample. It contains data fields:

- rental ID;
- scooter code (*deviceCode*) – reference to the list of scooters available for the region;
- status (*status*) – rental status that may have one of three values: *riding*, *pause*, *finished*. When scooter is used, it always has status “*riding*”. If the customer decides to stop, leave scooter on the street, and lock it for further use after some minutes, scooter is in status “*pause*”. These “*pause*” minutes should be counted because another tariff may be applied for this period;
- start time (*startTime*) – time when the scooter’s rental is started;
- start location (*startLocation*) – location where the rental is started;
- finish time (*finishTime*) – time when the rental is finished;
- final location (*finishLocation*) – final location of the scooter;
- pause minutes (*pauseMinutes*) – minutes spent for pauses;
- total distance (*totalDistance*) – total ride distance.

**C. Data Quality Runtime Verification Process**

Data quality runtime verification requires a business process definition, including states of process, possible events, and a data quality definition. The definition of the verification process for an example of the scooter rental process is shown in Fig. 3.

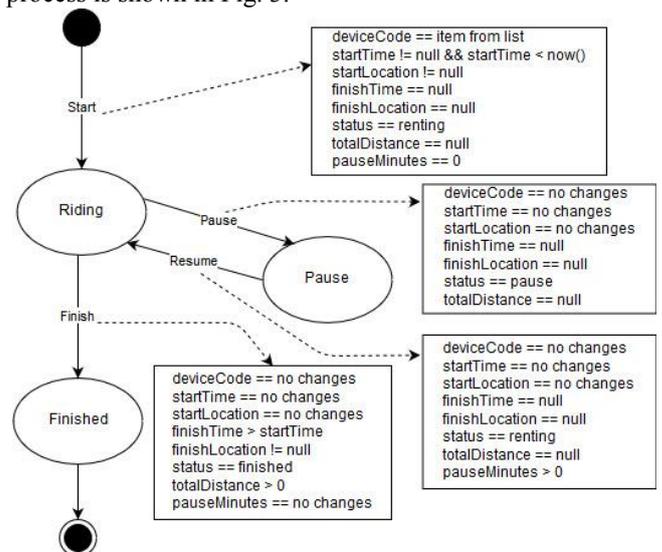


Fig. 3 Data quality verification of scooter rental process

When scooter rental starts, the initial data object should be extracted. According validation rules, the initial rental object contains information about rented device (reference to existing item from device list). *StartTime* should be different from null and less than *now()*, *startLocation* should be provided and the rental state should be “*renting*”.

After pause event is detected by runtime verification controller, a new data object version for verification should be extracted from database. New rules are applied to the object:

- *deviceCode*, *startTime* and *startLocation* remain unchanged (these values are set once when the object is created);
- *status* = “*pause*”;
- *pauseMinutes* should be unchanged comparing with the previous object version.

When riding activity is resumed (i.e., event “*resume*” is detected by the verification controller), the next copy of rental data object should be extracted, and a new set of rules should be applied:

- *deviceCode*, *startTime* and *startLocation* are unchanged;
- *status* = “*riding*”;
- *pauseMinutes* should be more than in the last version of object;
- *finishTime* and *finishLocation* are still null.

After ride finish event is executed and detected by runtime verification controller, the last version of rental data object is extracted from the business process database. The 4<sup>th</sup> set of data quality rules should be verified:

- *deviceCode*, *startTime* and *startLocation* are unchanged;
- *status* = “*finished*”;
- *pauseMinutes* are unchanged from the previous object version;
- *finishTime* and *finishLocation* are not null as the ride is finished and, moreover, *finishTime* > *startTime*.

As it can be noticed from the example, the runtime verification provides new possibilities for data quality verification:

- data quality may be applied and verified during business process execution and just for one object, not for the whole database;
- quality of data changes is verified by comparison of different versions of the same object;
- in the case of defect, the location of defect’s origin may be identified.

#### IV. ANALYSIS OF THE PROPOSED SOLUTION

The presented idea allows not only to ensure the process was correct and any logical or “external” constraint weren’t complied, but also to identify the moment and activity that caused or led to incorrect or inconsistent result.

As an example, let us imagine we have a database which quality we use to check once a month. We have already checked the quality of data of this database a month ago and it was of an excellent quality without any data quality issues or even anomalies. Now, a month later, we check it once again and find data quality problems not only in new records but also in those which were of good quality a month ago. It is difficult to detect the moment, when the data was changed, i.e. the qualitative data was replaced by data of poor quality, especially, if we don’t have access to log files where all activities are fixed. Moreover, in some cases such log files don’t exist at all or they are not detailed enough. However, the proposed idea of runtime verification in combination with data quality checks would solve this problem, detecting the moment and activity that caused the problem. To sum up, the main advantages of the proposed idea:

- data quality is verified immediately after the data is created/ modified;
- it is possible to detect the process step where the data is damaged;
- data quality evaluation is performed for the entire data set, not only for a specific data object that was changed;
- evaluation of the total data quality is reliable all the time;
- data verification can be performed independently of the system being executed.

However, there are also some potential disadvantages:

- if the runtime verification is performed incorrectly, it can lead to a tangible overload of the process being verified;
- by performing data runtime verification in parallel with system execution, data errors can be obtained for correct data if the process performs faster than the verification process and the data changes do not correspond to the step which data modifications are checked.

The proposed approach differs significantly from the Object Constraint Language (OCL) approach, which is designed to protect the database against incorrect value input but does not detect errors in input data.

As for the proposed solution, the main scope of data quality checks is data object retrieval. Data for their further quality checking are retrieved from agents by using denormalization as it significantly speeds up data retrieval. The denormalization must be implemented dynamically without knowing the denormalized relational target structure in advance [10].

The quality checking analysis runs in accordance with the initial runtime verification mechanism. In scope of the one separate check for the business process to be analysed, it can be compared with the model checking and testing mechanism using pre- and post- conditions proposed in [11]. By precondition is meant the result of previous check that is used as an input, however postcondition checking is

performed when an execution of verification completes. It is obvious that preconditions should be correct to be suitable for usage in quality checks. In our case, all preconditions are considered as correct as they are analysed at the previous stages/ steps. However, another assumption is that the initial data at the initial state of a check (let call it  $q_0$ ), that also is a precondition, is also correct as any data modification was done previously, as it is assumed that statically stored data (that isn't involved in any process) is checked from time to time and as a result is correct.

The frequency and number of data quality checks as well as points when they should be done depend on the use-case and user's preferences. This idea corresponds with the data object-driven approach to data quality evaluation allowing users to take control over every step of data quality analysis process. As a result, the proposed solution respects quality checks: (a) after every step as well as (b) only when the user considers them as important, or (c) with periodical frequency after a particular step, for instance, once a day or every time after specific process is finished etc. The first option ensures in-depth and comprehensive quality analysis, when every step is checked. However, as there might be cases, when several steps are not of high importance at least for a particular user, or there is no necessity in continuous checks, for instance, in order to save resources and efficiency, the second and third options appear suitable. Moreover, in the future, the idea of prioritization mechanism would be evaluated to provide users the possibility to perform some checks with higher priority first. This mechanism would offer to users a higher level of control over the whole process.

The number of cases when the proposed solution can be useful is high, including the continuous assessment of the quality of information systems, e-government [12], etc. by data quality runtime verification. The proposed solution can lead to the improvement of quality of many services increasing government effectiveness and quality of public services [13] that nowadays become a topical issue.

## V. CONCLUSION

This paper deals with runtime verification for checking data quality during business process execution. The user defines data objects and requirements of his/ her interest using graphic DSLs and ensuring high quality of data object parameter values.

Unlike other data quality studies, the proposed solution provides an operational runtime verification of data quality requirements, allowing to detect deviations from quality requirements in the particular business process's execution step when incorrect data object parameters are recorded in the database. Verification of data quality requirements and the base process are parallel processes. Impact of verification on the base process performance is insignificant and acceptable if sufficient hardware capacity is available.

The proposed solution is an "external" solution that checks the data quality requirements without direct connection to the business process. Such approach allows enabling and disabling of runtime verification of data quality requirements operationally at various stages of information system usage. Graphical DSLs that is used to describe data objects and quality requirements is intuitively understood and suited for use by non-IT and data quality specialists.

In the future, the proposed approach might be applied to issues of the semantic web.

## REFERENCES

- [1] El Hadji Bassirou Toure, I. Fall, A. Bah, M. S. Camara, *Megamodel-based Management of Dynamic Tool Integration in Complex Software Systems*. In FedCSIS Position Papers, 2016, pp. 211-218, <http://dx.doi.org/10.15439/2016F585>.
- [2] I. Oditis, J. Bicevskis, *Asynchronous Runtime Verification of Business Processes: Proof of Concept*. International Journal of Simulation-Systems, Science & Technology, 2015, 16(6), 1-11, <http://dx.doi.org/10.5013/IJSSST.a.16.06.06>.
- [3] J. Bicevskis, Z. Bicevska, A. Nikiforova, I. Oditis, *An Approach to Data Quality Evaluation*. In 2018 Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS), IEEE, 2018, pp. 196-201, <http://dx.doi.org/10.1109/SNAMS.2018.8554915>.
- [4] J. Bicevskis, Z. Bicevska, A. Nikiforova, I. Oditis, *Data quality evaluation: a comparative analysis of company registers' open data in four European countries*. In Communication Papers of the Federated Conference on Computer Science and Information Systems (FedCSIS), 2018, pp. 197-204, <http://dx.doi.org/10.15439/2018F92>.
- [5] A. Coronato, A. Testa, *Approaches of Wireless sensor network dependability assessment*. In 2013 Federated Conference on Computer Science and Information Systems, IEEE, 2013, pp. 881-888.
- [6] A. Nikiforova, *Open Data Quality Evaluation: A Comparative Analysis of Open Data in Latvia*. Baltic Journal of Modern Computing, 2018, 6(4), 363-386, <https://doi.org/10.22364/bjmc.2018.6.4.04>.
- [7] A. Nikiforova, J. Bicevskis, *An Extended Data Object-driven Approach to Data Quality Evaluation: Contextual Data Quality Analysis*. In Proceedings of the 21st International Conference on Enterprise Information Systems - Volume 1: ICEIS, 274-281, 2019, <http://dx.doi.org/10.5220/0007838602740281>.
- [8] A. Nikiforova, *Analysis of Open Health Data Quality Using Data Object-Driven Approach to Data Quality Evaluation: Insights from a Latvian Context*. In IADIS International Conference e-Health 2019, Part of the IADIS Multi Conference on Computer Science and Information Systems, MCCSIS 2019, IADIS
- [9] G. C. Deka, *NoSQL: database for storage and retrieval of data in cloud*, Ed. CRC Press, 2017, <https://doi.org/10.1201/9781315155579>.
- [10] C. Gröger, F. Niedermann, B. Mitschang. *Data mining-driven manufacturing process optimization*. In Proceedings of the world congress on engineering, 2012, Vol. 3, pp. 4-6.
- [11] S. Khurshid, C. S. Păsăreanu, W. Visser, *Generalized symbolic execution for model checking and testing*. In International Conference on Tools and Algorithms for the Construction and Analysis of Systems. Springer, Berlin, Heidelberg, 2003, pp. 553-568, [https://doi.org/10.1007/3-540-36577-X\\_40](https://doi.org/10.1007/3-540-36577-X_40).
- [12] E. Ziembra, T. Papaj, D. Descours, *Assessing the quality of e-government portals-the Polish experience*. In 2014 Federated Conference on Computer Science and Information Systems, IEEE, 2014, pp. 1259-1267, <http://dx.doi.org/10.15439/2014F121>.
- [13] A. Karabegovic, M. Ponjavic, *Geoportals as decision support system with spatial data warehouse*. In 2012 Federated Conference on Computer Science and Information Systems (FedCSIS), IEEE, 2012, pp. 915-918, <http://dx.doi.org/10.13140/RG.2.2.26385.68963>.



# BPM Tools for Asset Management in Renewable Energy Power Plants

Carchiolo Vincenza\*, Catalano Giovanni<sup>†</sup> Malgeri Michele<sup>†</sup>, Pellegrino Carlo<sup>†</sup>, Platania Giulio<sup>†</sup>, Trapani Natalia<sup>†</sup>,

\*Dipartimento di Matematica ed Informatica - Università di Catania - Catania - Italy

<sup>†</sup>Dipartimento di Ingegneria Elettrica Elettronica Informatica - Università di Catania - Catania - Italy

<sup>‡</sup> Development and Support Center - BaxEnergy - Catania, Italy

**Abstract**—Business Process Management (BPM) is an accepted discipline and its importance in increasing automation inside industrial environment is today recognized by all players. The complexity of modern management process will lead to chaos without a well-designed and effective BPM. Several BPM Suites were compared and BPM approach was applied to the case study of process management in a renewable energy power plant. Results both in process reduction and simplification and flow optimization obtained in the real case are discussed to state efficacy and efficiency of the adopted approach.

## I. INTRODUCTION

**I**N A competitive environment in which companies have to provide more and more effective and efficient services/products, asset management allows to obtain value from assets and achieve the company's business objective [1]. To accomplish both the goals to ensure good operational performance and long durability of the final products/services, it is necessary to define effective business processes, to monitor their performances and to provide corrective actions when necessary. In fact, the quality of the operational processes are more and more dependent on maintenance processes, thus they must be carefully engineered and effectively implemented. Maintenance management changed in the last decades thanks to ICT development, from a management perspective, Computerized Maintenance Management System (CMMS) has contributed to enhance the control of maintenance activities [2] and maintenance is considered a relevant business function, able to interact with all other strategic functions, such as operations, purchasing, suppliers (service providers), warehouses management, administration, therefore maintenance effectiveness and efficiency is crucial for business.

BPM frameworks provide tools that allow to *design* the business process using model, maps and rules, than it allows to implement them defining the architecture and adding rules. Moreover, framework executes and monitors the process (providing processes measurement) collects data, and, finally, it provides analysis and diagnosis that allows to improve the process itself [3].

Indeed, the process can be viewed from different perspectives thus producing multiple models and styles [4]: *control flow perspective*, *resource perspective* that means to focus on equipment, units, etc.; *data creation perspective*; *time perspective* that focuses on deadlines and *function perspective* that uses activities.

BPMN 2.0 is a logical description of business processes and how they operate that focuses on process implementation and Process simulation [5]. Usually, the business process simulation collects several data and provide a visual animation of the evolution of the process. The analysis of data permits the designer together with supervisor and manager to identify (potentially expensive) mistakes and, thanks to the Key Performance Indicators (KPIs), defined during the early designing of the process, increase the performance to getting and overall improvement of the process (e.g. removing bottlenecks).

In literature, the use of the BPM approach for maintenance management is not so common [6] (see references in [7], [8]). Most of them suggest that an Asset Management System Software can help organizations to achieve operational excellence, through a more effective cost control, a more efficient asset planning, a reducing in capital expenses, an optimization of operational costs, thus extending asset life-cycle and obtaining a higher Return On Asset.

This works briefly presents some of the criteria that were adopted by BaxEnergy© to develop a BPMS, starting from existing platform aiming at optimizing the maintenance of *renewable energy power plants* to be offered as a service to different companies. Renewable-energy Power Plant faces different challenges [9], [10], [11]. Finally, a case study is discussed highlighting the enhancement and discussing some details that lead to final implementation.

Section II introduces BPMS and some of the criteria used to compare the platforms, and provide the reader with a comparison among several platforms that was available at the moment of the study. Section III discussed the case study developed using the selected platform.

## II. BPMS COMMON FEATURES AND CRITERIA

BPMN can be useful both for planning complex business processes and to control and monitoring them once deployed. A business process can be defined as a set of connected activities that create value for customers; usually, they are classified into three groups: *strategic*, *operational* and *support*. A BPMN provides users and developers with tools that clearly denote the objective, who is in charge of any operation, starting and ending points, input and outputs, constraints and monitoring points. One of the most important aspect of BPMN is surely the graphical model that, usually, is formally

documented thanks to UML (Unified Modelling Language) class diagrams [5].

BPMN 2.0 is the current standard that adds the following characteristics: *human-readability*, that is a standard visual notation for modelling processes; *accessibility*, that means that all actors must understand what is represented; *machine-readability*, that implies the use of the XML notation for simulating and executing processes [12].

BPMN Graphical representations use several standard symbols that must be easily understood by both developers and users the most important are: *swim lane*, *pool*, *box*, *event*, *task*.

A fundamental element to make dynamic the process diagram is *gateway*: it allow the designer to choose where the flow of a diagram must follow a path or another by evaluating a condition. Two types are defined in the standard: *XOR* or *Exclusive*. Gateway is the task that drives the evolution of a process by evaluating a condition and selecting the next task, *OR* or *Parallel* Gateway split the evolution into two parallel branches that must be joined with the same element when the evolution of the process does not support parallel activities.

#### A. Evaluation Criteria

There is no doubt that some interesting features for a Software developer are meaningless by the point of view of a Business Process developer, and are meaningless for a standard employee, therefore the evaluation Criteria are grouped into two classes. The first group deals with the feature useful to software developers, the most important are as follows:

- *OpenSource*, software suite refers to open source model, that means that products are released under an open-source license, this feature allows to inspect the source and to add any additional features;
- *Community*, the community supporting the software suite should be as large as possible to share problems and their solutions. A large community implies Documentation and Tutorial are easy to find and this is useful both for software and model developers;
- *BPMN 2.0 supported*, this is a must for any new software, since today BPMN 2.0 standard is largely supported;
- *Additional modules and connectors*, they permits to extend the functionality of the Software and to write new custom connectors;
- *API provided*, this can be useful to easily add functionalities to the BPMS.
- *Innovation*, of course, it is important that software solution supports all recent technologies and methodologies but it must be also "mature" (according, for instance, with Gartner Magic Quadrant [7]). Some examples of innovative features are mobile device deployment without the need to develop a dedicated application, low-code Platforms, integration with Artificial Intelligence for performance analytics and so on;

Second group of features, that we could call *Functionality-based evaluation criteria*, belongs to Business Process Developers that focus on the processes themselves and, usually, have

no specific experience in software development. They need to model, describe and test production processes aiming at optimizing them according to some metrics (e.g. cost, duration, response time) often conflicting. The most important are listed below:

- *Web Modeler/Collaboration enabled*, it enables collaboration between developers;
- *Template Library*, the presence of a library of reusable models is a valuable add-on to any Suite;
- *Model and Process Versioning*, since processes run for long times, the ability to control the current version of Models and the ability to roll back to a previous version are very useful;
- *Powerful graphic interface*, the interface should support advanced interface functions as, for instance, Drag&Drop of models/processes, Form Editing that allows final users to create and edit data, to customize the colors of element in a diagram to make it more readable;
- *Process deployment*, the system should allows to make an instance of the modelled process to check errors and to locate performance indicators or failures;
- *Testing and Simulation*, the presence of an engine for simulating the process and/or validate the model is an essential feature;
- *Customizable Properties*, that means the ability to change the properties of the trial modelling tool.
- *Business rule engine and activity Monitoring*, possibility to integrate business rules for the process and to monitor the execution of a process;
- *Integration with Cloud*: since some BPMS are integrated with cloud service, it is important to support the integration between BPMS and Cloud.
- *Role Based Security*, that allows to manage of the security rules for each role.

#### B. A Brief Survey of BPMNs Platforms

The evaluation criteria used by throughout this work are based on Critical Capabilities such as Interaction Management, Monitoring and Business Alignment, Rules and Decision Management, Analytic and they are evaluated on the base of some Use Cases. Great importance is given to Continuous Process Improvement (referred as CPI) and to Citizen Developer Application Composition (that is the ability to leave aside from IT development staff) [5].

The BPM suites, that was evaluated, support highly intelligent applications which integrate more-advanced decision automation technologies (e.g. predictive analytic, artificial intelligence) and decision support for knowledge workers to automate business processes that require an adaptive behaviour [5]. It means that actually *iBPMS* is not a simple modelling software or a validation engine for BPMN, but it is both bound and integrated with other functionalities that keeps track of statistics and, sometimes, let even the user to define performance indicators.

The main characteristics of the tools were summarized and compared in tables I and II, the former deals with features

useful to developers and the latter to features useful to users, i.e. the provided functionalities.

TABLE I  
CRITERIA RELATED TO DEVELOPMENT

	OpenSorce	Community	Innovation	API
BPMN.IO	Y	-	Y	-
Cawemo	-	-	Y	-
Camunda	Y	Y	Y	Y
Bizagi	-	Y	Y	Y
Appian	-	Y	Y	Y
Activiti	Y	Y	Y	Y
IBM	-	Y	Y	Y
jBPM	Y	Y	Y	Y
KissFlow	-	Y	Y	Y
QuickFlow	-	-	-	-

TABLE II  
CRITERIA RELATED TO FUNCTIONALITY

	Web Functions	Versioning	Testing	Monitoring
BPMN.IO	P (Partial)	-	-	-
Cawemo	P	-	-	-
Camunda	Y	Y	P	Y
Bizagi	Y	Y	Y	Y
Appian	Y	Y	Y	Y
Activiti	Y	Y	-	Y
IBM	Y	Y	Y	Y
jBPM	Y	Y	Y	Y
KissFlow	Y	-	-	-
QuickFlow	Y	-	-	(SalesForce)

Since the use of opensource platform covering all the process, from development to monitoring, exposing a flexible ReST API to integrate easier with existing software are the most important requisites the selected platform was Camunda. The lack of processes Simulation is overcome thanks to the a great support of Javascript libraries that can be integrated with no effort.

### III. CASE STUDY

#### A. The Company

The case study was developed, during the research project WEAMS, in an Austrian company that is a BaxEnergy© client. The Company's core business is the construction and operations management of wind turbine power plants. The Company produces electrical energy through wind turbines, by itself or cooperating with other partners or investors. In the case study was considered that it actually deals with operations and maintenance processes of wind turbine farms.

After studying and analyzing their general organizational structure, we find 92 different relevant processes. These processes (e.g., tool provisioning, maintenance execution in harsh weather conditions, invoicing, checking financial guarantees, executing training, and so on) include several tasks with their relevant costs. In general we identified two main process categories related to wind turbines maintenance: *Core processes* (usually technical ones) and *Support processes* (usually organizational and financial ones).

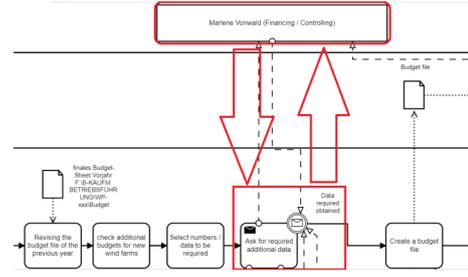


Fig. 1. Collaboration between actors of a process

#### B. Process Re-engineering

A subset of 39 processes out of the 92 relevant for the company were selected as they are the most frequently activated during a year and relevant also for other companies in the same business sector. Moreover, other ten maintenance core-processes, not previously engineered by the company were identified and implemented in order to guarantee better maintenance performance.

A first revision of the processes allowed to increase the general comprehension of process workflows, by the Company. In particular, to follow the correct logic of BPMN 2.0:

- The use of *send* and *receive* task was corrected;
- Some *End event* were added to complete some workflows;
- Some *Intermediate* events instead of useless tasks were introduced in order to simplify some process;
- Some *script tasks* were modified into *normal tasks* (because scripts will be added in a future phase);
- The use of *gateway* was corrected and useless ones were eliminated.

Furthermore, some specific processes were deleted as *stand-alone processes* and they were incorporated within other ones, by producing a unique layout representing the whole activity without missing any essential information and better visualizing interactions through inputs and outputs.

This *process synthesis* generated a total incorporation of 7 process into other ones, thus further reducing to 26 processes out of 39.

Let us note, the presence of a particular interaction between pools and collapsed pools inside the processes. Those interactions called *swim lanes* in the BPMN 2.0 standard are used mostly to enable the collaboration between core and supporting parts of the processes.

For example, in the process *Budgeting* a supervisor (financing/controlling) of the process needs to receive some additional information before going on with next task. In this case, the supervisor has to do two actions (see Figure 1): Before the re-engineering, the company used the above described interaction also to communications with the ticket system of WEAMS. Thanks to the BPM system, implemented in the WEAMS system, this use of send/receive tasks is no more necessary (see Figure 2) but it requires just a simple task in WEAMS system (see Figure 3). In this way, there is a

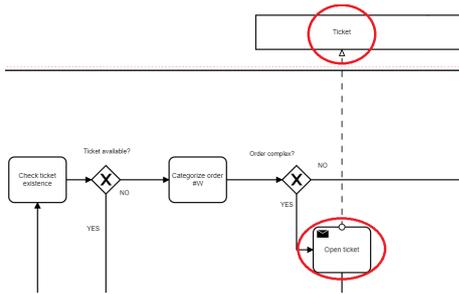


Fig. 2. Collaboration with the ticketing system of WEAMS

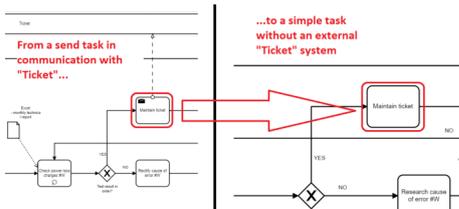


Fig. 3. From ticket system to a simple task

reduction of about 70 message flows due to ticket system implemented on the BPM system in WEAMS.

These relevant 26 processes were subjected to a deep revision in order to optimize elements and workflows and simplify them. Such a revision allowed to eliminate from the processes 26 phases and 106 transitions that was useless, obtaining more efficacy and efficiency in process workflow and improving the general performances of the maintenance processes.

In the second step of revision some of the 26 processes were splitted into two or more sub-processes in order to easily implement them into the WEAMS BPM system, thus generating 31 processes.

In the third step some maintenance processes were added to manage some aspects that are relevant for the Company business but that never were implemented by the Company, such as Planning ordinary maintenance's intervention, Human resources management, Warehouse management with spare parts and consumables, Personnel qualification management to meet normative requirements, Corrective maintenance management (ticket opening, see Figure 4), Predictive maintenance management (specifically inspections, see Figure 5), Asset status management, Service level management, Maintenance on the field and Logistics management.

### C. Key Performance Indicators

The section in which KPIs have the biggest role is the wind turbine maintenance and the most influencing factors are certainly economical, technical and organizational. Maintenance performance indicators reflect achievement and progresses in meeting a goal; clearly, the greater is the installed capacity the higher are Operational and Maintenance costs.

Leading indicators measure performance before the maintenance process results starts to follow a particular trend and

monitor if maintenance activities are producing good results in a long-term period. An example *Preventive Maintenance (PM) Completion Rate*: a low completion rate for PM would generate an increasing in asset maintenance work while a high completion rate means that asset preventive maintenance request is correctly being completed and, probably, future corrective maintenance requests will be reduced. Another example is the *Outage Schedule Compliance* an important metric to track future maintenance work because it allows to measure deferred asset maintenance, resulting in an increased risks and likelihood that asset performance will decrease at a future time, leading to lower capacity, increased downtime, and greater expenses.

Lagging indicators use historic data to obtain a measure to confirm coherence with long-term performance trends; they are used to determine how well a process performs.

In order to increase maintenance performance, both internal and external factors of a company should be considered as complex activities. Therefore, considering each influencing factor is essential to assess, control, measure and compare performances. The KPIs in technical standards (specifically UNI EN 15341: 2007) can be grouped into three categories: *economical* (E, 21 KPIs), *technical* (T, 21 KPIs) and *organizational* (O, 26 KPIs). A selection of them was considered to be linked to the maintenance related processes implemented as shown in Table III. These allow the Company to calculate performances of related processes.

## IV. CONCLUSIONS AND FUTURE WORK

Introducing BPMN 2.0 into an asset management model is an efficient way for controlling and sharing information between all actors involved in a process. Specifically, in maintenance management of wind turbines there are a lot of factors (e.g. weather conditions, personnel availability) and events such as logistics, administrative or technical ones (e.g. failure) that can change in an unpredictable way the process performances. The developed WEAMS BPM system allows the definition and execution of management processes within renewable energy power plant, simplifying relations among company functions, introducing standard activities for each process, assigning tasks to users, eliminating non added value phases thus providing an overall reduction of downtime of wind turbines, procurement optimisation due to higher efficiency in warehousing, human resources management, maintenance cost reduction.

### ACKNOWLEDGMENT

This work was partially supported by WEAMS N. F/050145/00/X32 project.

### REFERENCES

- [1] The Institute of Asset Management, "IAM asset management maturity guide v1.1," The Institute of Asset Management, Tech. Rep., Jun 2016. [Online]. Available: <http://www.theiam.org/>
- [2] N. Trapani, M. Macchi, and L. Fumagalli, "Risk driven engineering of prognostics and health management systems in manufacturing," *IFAC-PapersOnLine*, vol. 48, no. 3, pp. 995 – 1000, 2015, doi: 10.1016/j.ifacol.2015.06.213.

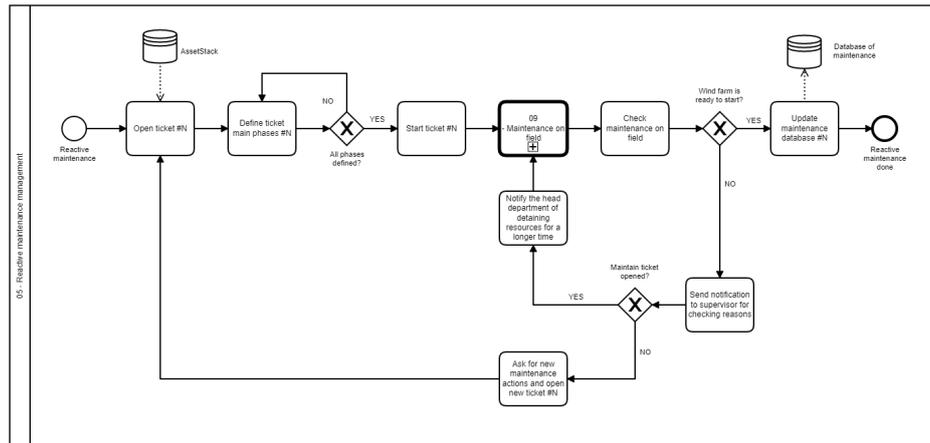


Fig. 4. Corrective Maintenance Management Process

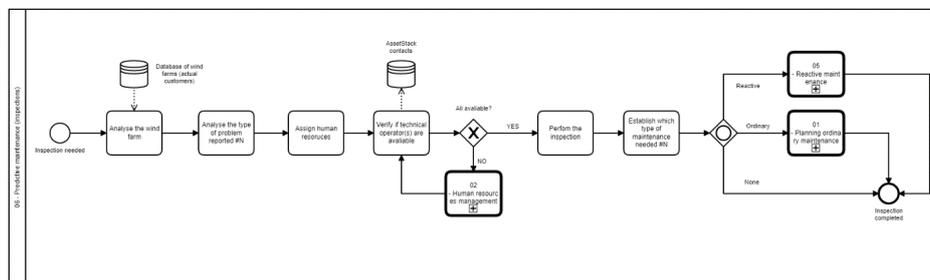


Fig. 5. Predictive Maintenance Management Process (inspection)

TABLE III  
SELECTED KPIS & RELATED PROCESSES

Type	Factors	Related processes
E1	Total maintenance costs / Asset, replacement value	01; 02; 03; 04; 05; 06; 07; 09; 10
E3	Total maintenance cost / Quantity of output	01; 02; 03; 05; 06; 07; 08; 09; 10
T1	Total Operating time / Total Operating time + Maintenance Downtime	01; 05; 06; 07; 08; 09
T8	Preventive maintenance time causing downtime / Maintenance Total downtime	06; 09
O1	Number of internal maintenance personnel / Total internal employees	01; 02; 05; 06; 09
O2	Number of indirect maintenance personnel / Number of internal maintenance personnel	01; 02; 05; 06; 09
O3	Number of indirect maintenance personnel / Number of direct maintenance personnel	01; 02; 05; 06; 09
O16	Corrective maintenance man-hours / Total maintenance man-hours	01; 05; 06; 09
O18	Preventive maintenance man-hours / Total maintenance man-hours	01; 05; 06; 09
O22	Work orders performed as scheduled / Total scheduled work orders	01; 02; 05; 09

[3] J. Montilva, J. Barrios, I. Besembel, and W. Montilva, "A business process model for it management based on enterprise architecture." *CLEJ online vol.17, n.2*, pp. 4–4, 08 2014, doi: 10.19153/cleiej.17.2.3.

[4] W. M. van der Aalst, M. La Rosa, and F. M. Santoro, "Business process management. don't forget to improve the process!" *Business & Information Systems Engineering 58(1)*, pp. 1–6, 01 2016, doi: 10.1007/s12599-015-0409-x.

[5] T. Allweyer, *BPMN 2.0: Introduction to the Standard for Business Process Modeling*. Books on Demand, 2016. [Online]. Available: <https://books.google.it/books?id=sowaDAAAQBAJ>

[6] M. Jasiulewicz-Kaczmarek, R. Waszkowski, M. Piechowski, and R. Wyczolkowski, "Implementing BPMN in maintenance process modeling," *Advances in Intelligent Systems and Computing*, vol. 656, pp. 303–309, 01 2018, doi: 10.1007/978-3-319-67229-8\_27.

[7] Gartner, "Magic quadrant for intelligent business process management suites," Gartner, Tech. Rep., 2019, accessed 10 May 2019. [Online]. Available: <https://www.gartner.com/en/documents/3899484>

[8] I. Corporation, "Understanding the impact and value of enterprise asset management." <https://www.ibm.com/downloads/cas/XJRD7M1Z>, ©IBM Corporation, Tech. Rep., 2016, accessed 10 May 2019.

[9] Accenture, "The future of onshore wind operations and maintenance," Accenture, Tech. Rep., 2017, accessed 10 May 2019. [Online]. Available: <https://www.accenture.com/us-en/insight-future-onshore-wind-operations-maintenance>

[10] M. Shafiee and J. D. Sørensen, "Maintenance optimization and inspection planning of wind energy assets: Models, methods and strategies." *Reliability Engineering and System Safety*, pp. 1–19, 2017, doi: 10.1016/j.res.2017.10.025.

[11] J. Wang, X. Zhao, and X. Guo, "Optimizing wind turbine's maintenance policies under performance-based contract." *Renewable Energy, Volume 135*, pp. 626–634, 05 2019, doi: 10.1016/j.renene.2018.12.006.

[12] A. Ciaramella, M. G. Cimino, B. Lazzarini, and F. Marcelloni, "Using bpmn and tracing for rapid business process prototyping environments." in *ICEIS 2009 - 11th International Conference on Enterprise Information Systems, Proceedings*, 01 2009, pp. 206–212, doi: 10.5220/0002005002060212.



# Identification of Heuristics for Assessing the Usability of Websites of Public Administration Units

Łukasz Krawiec, Helena Dudycz  
Wrocław University of Economics, Wrocław, Poland

Email: {lukasz.krawiec, helena.dudycz}@ue.wroc.pl

**Abstract**—A very important aspect of modern websites is their usability. Thanks to modern, and constantly developing technologies it is possible to create user-friendly services for each user. The usefulness of online services may be considered in terms of their functionality, clarity, and accessibility. It is particularly important that these criteria are met by public administration websites. The aim of this paper is to present the most common usability errors identified on the websites of public administration units as well as to indicate the links between particular types of problems and traditional heuristics of Jakob Nielsen. The survey was conducted by evaluating the websites of the Public Information Bulletin in Poland (BIP, which stands for “Biuletyn Informacji Publicznej”), which are supposed to provide universal access to public information by the citizens of the country. A heuristic method (based on J. Nielsen's heuristics) was used to evaluate 60 websites. The errors obtained were grouped into 14 categories. Each of the error groups was assessed by an expert in terms of its importance for the overall assessment of the website's usefulness. The analysis of relations between the identified categories of errors and heuristics of Jakob Nielsen indicates a need for specifying heuristics in the context of evaluating the usability and availability of public administration websites.

## I. INTRODUCTION

Regardless of the motives of designers or clients – the purpose of websites mainly comes down to the effective presentation of their content and efficient conveyance of information (usefulness) to the largest possible audience (accessibility). This means that it is important to ensure that both healthy and disabled people are able to effectively familiarise themselves with the information provided on the website and take advantage of its functionalities.

As far as creating useful and accessible websites of public administration units is concerned, it is necessary to conduct continuous research and usability tests, perceiving it as one of the basic activities in the process of developing such websites. The aim of such activities is to prevent dissatisfaction among users (i.e. citizens) and to provide a place where they can find the information they need quickly and efficiently.

The aim of this paper is to categorise the most common errors identified on the websites of public administration units as well as to indicate the links between particular types of problems and traditional heuristics of Jakob Nielsen. This

will help adjust the heuristic method to the needs of further usability studies of this kind of websites.

The structure of the paper is outlined below. The next section briefly explains the concept of usability in the context of websites. The section that follows focuses on the characteristics of the heuristic method for testing and evaluation of usefulness. Next, the Polish Public Information Bulletin is briefly described. The penultimate section presents the proposed procedure for the examination and the results obtained. Finally, a summary of the paper is provided.

## II. THE USEFULNESS OF A WEBSITE

In literature, usability is defined in a variety of ways. According to ISO 9241 [1], usability defines “the extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use”, while the standard ISO / IEC 9126-1 (for Standardization and Commission, 2001), related to Software Engineering and product quality, describes usability as the ability of the software product to be understood, its operation learned, to be operated, and to be attractive to the user. In the literature, usability is defined as the “capacity to be used” the device [2] and depends on what the user wants to do [3].

According to J. Nielsen, usability “is a quality attribute that assesses how easy user interfaces are to use” [4], comprising 5 components:

- Learnability: How easy is it for users to accomplish basic tasks the first time they encounter the design?
- Efficiency: Once users have learned the design, how quickly can they perform tasks?
- Memorability: When users return to the design after a period of not using it, how easily can they re-establish proficiency?
- Errors: How many errors do users make, how severe are these errors, and how easily can they recover from the errors?
- Satisfaction: How pleasant is it to use the design? [4].

The studies described in the literature [5] indicate that usability is the most important parameter affecting the quality of websites evaluated by their users. According to Paplauskaitė [6], the usability of the website determines its legibility, intuitiveness, and comfort of use. The concept

of web usability is connected with the concept of web accessibility. It means that people with disabilities have full access to the content of a given website, can understand it as well as benefiting from convenient navigation and interaction with the website [7]. It can, therefore, be concluded that the accessibility of websites is related to the human-computer interaction and is a feature of the user interface that allows all people to use it, regardless of their hardware, software or disability. Accessibility problems are most common among users with reduced mobility, hearing or vision, including those with cognitive disorders [8, p. 41; 9, p. 169]. Accessibility is now seen more broadly, i.e. the aim is to make the website accessible to as many people as possible, including the elderly, people with disabilities, people with low bandwidth internet access, and people using older devices, which are usually slower than modern ones [10].

P. Morville names usability and accessibility as two separate dimensions out of the six that make up User Experience, altogether creating a profit or value for the user, ensuring that they receive a product that meets their needs [11]. According to other authors, usability is a broader concept, a subset of which is accessibility, including issues such as interface handling problems experienced by people with disabilities [12, p. 7; 13].

### III. THE HEURISTIC METHOD APPLIED FOR EXAMINING WEBSITES

The literature discusses many ways to study the usefulness of websites [2; 14-18]. One of them is the heuristic method, which is one of the expert, inspection-based techniques of recognising usability problems. It consists in indicating the extent to which a given piece of software or a website complies with the developed rules and standards (called usability heuristics [2]) for the design of human-computer interactions. In this method, experts indicate what is correct and what is incorrect about the website being evaluated in terms of the heuristics applied [19].

The heuristic analysis of a website is a universal and easily applicable method. It is used for researching entire websites, as well as only one or two pages of a given website. It is a relatively inexpensive method as there is no need to involve users and the indicated number of experts is limited, i.e. three to eight experts are considered to be the optimal number [20]. An independent analysis performed by each of them supports the study's objectivity of and effectiveness. Thanks to this method, it is possible to detect many small as well as major errors related to the website's performance. Also, it allows one to identify the elements of the website that may adversely affect its usability.

The literature most often refers to heuristics developed by J. Nielsen, also referred to as traditional ones. These are [4]:

**H01. Visibility of system status** The purpose of the system is to inform the user about what is currently happening while

working with the system, e.g. by sending a message in situations where the system's response time is longer than usual or by placing very helpful progress bars while the user is performing a process consisting of several steps. Sounds or backlighting can be also applied to enhance feedback.

**H02. Correspondence between the system and the real world** The system should avoid technical terms and use only terms and expressions known to the user. In addition, the system should present only the information that is actually needed by the user. It is recommended that it be naturally and logically ordered.

**H03. User control and freedom** The result of the user's actions within the system should be reversible. It often happens that a person using a certain solution mistakenly chooses an option other than the desired one. In such a situation, the system is required to be able to revert activities without having to go through successive stages of the process with an incorrectly selected variant or repeat all the steps from scratch. A well-designed system should allow the user to pause their activities at any time and resume in the same place after the interruption.

**H04. Consistency and standards** The system should be consistent visually (the appearance of windows, colours, the layout of buttons, etc.), operationally (same way of starting operations, same keyboard shortcuts, etc.), and behaviourally (the system's expected reaction to the user's actions). The person using the product should have no doubt about whether similar phrases or actions always mean the same. It is recommended that the conventions applied to the whole platform be adhered to.

**H05. Prevention of errors** It is recommended that situations in which human error is likely should be detected and removed. As for uncertain situations, the system should ask the user if they are sure whether they want to execute the command. It is also worth using various forms of facilitation that will effectively eliminate common errors, for example, by checking spelling, grammar, or command line correctness.

**H06. Recognition and not remembering** The system should not require the user to remember information between successive stages of the dialogue. Access to information relating to the operation of the system should be possible from any location. In addition, it is important to ensure that all available options and actions are clearly visible.

**H07. Flexibility and efficiency of use** A desirable feature of a good system is that it allows operation using shortcuts. Activities often performed by the user should be flexible and adaptable to their needs. It is recommended that keyboard shortcuts, auto-supplements, lists of recently used commands, quick access bars, etc. be used. For tedious, multi-step processes, the system should allow the user to create macro commands.

**H08. Aesthetic and minimalist design** It is recommended to avoid placing unnecessary and distracting elements in the dialogue. They reduce the focus of the person using the solution on the proper content of the task. It should also

be remembered that simple designs, with a small number of elements, indicate the system's ease of use.

**H09. Help users recognise, diagnose, and recover from errors** Error messages sent by the system should be written in a language that is easy for the user to understand. They should carefully explain the cause of the error and suggest a way to repair it.

**H10. Help and documentation** It is recommended that every system have access to a user manual that should not be too extensive. The user, using this type of manual, should be able to easily find the information they need. Well-designed manuals describe the steps the user needs to take to restore their system back to normal.

In addition to the heuristics described above, the literature provides many other approaches to evaluating usefulness with this method, including the following:

- Cognitive Engineering Principles for Enhancing Human-Computer Performance [21],
- Weinschenk and Barker classification [22],
- The Eight Golden Rules of Interface Design [23],
- Usability Heuristics for Touchscreen-based Mobile Devices [24],
- First Principles of Interaction Design [25],
- 7 Usability Heuristics That All UI Designers Should Know [26].

Many of the above rules and guidelines are based on J. Nielsen's classic heuristics. The aim of many heuristics creators is to update and match them to the study of specific IT systems [27-28]. New heuristics proposals also result from a change in the way of looking at the interface usability issue. For example, the aforementioned Gerhardt-Powals [21] takes a more holistic approach to evaluation, including principles such as: automate unwanted workload, group data in consistently meaningful ways, practice judicious redundancy. A more detailed and fragmented approach is proposed by Susan Weinschenk and Dean Barker [22] on their list of twenty guidelines. These are among others: user control, accommodation, simplicity or predictability. Ben Shneiderman's goal was to create flexible principles that can be adapted to interfaces in different programming environments. For example: strive for consistency, seek universal usability, permit easy reversal of actions [23]. A similar point of view is represented by Bruce Tognazzini's guidelines, such as: aesthetic design, anticipation, autonomy, discoverability [25]. Most interface usability experts follow similar principles or build on existing proposals.

#### IV. THE PUBLIC INFORMATION BULLETIN AS AN EXAMPLE OF A PUBLIC ADMINISTRATION WEBSITE

From the perspective of the public interest, especially in the age of the information society, all public administration websites should offer features such as usability, or accessibility, which is inherent in it. The most important and widespread standard for this feature in the world is the WCAG (Web Content Accessibility Guidelines). Many countries are implementing additional

recommendations and legal requirements to ensure the quality of public websites containing information and content of particular interest to the general public. Among the examples thereof are the US Section 508 of the Workforce Rehabilitation Act [29], the German Barrierefreie-Informationstechnik-Verordnung [30], or the Italian Stanca Act [31], adjusting the law to the W3C WCAG 2.0 accessibility requirements.

In Poland, the Public Information Bulletin (BIP) is an example of a website of public administration units, constituting a unified system of Internet services ensuring free-of-charge and universal access to public information in Poland. Apart from the main website of the Public Information Bulletin (<https://bip.gov.pl>), the bulletin consists of services provided by entities obliged to maintain them, such as public authorities, economic and professional self-government bodies, entities representing state organisational units, political parties, and many others. Their task is to inform the public about their activity, i.e. to make public information available. Additionally, detailed requirements and recommendations for BIP administrators can be found on the website of the Ministry of Digitisation (<https://bip.gov.pl>). The straight majority of the above quality requirements come down to the concept of usability. BIP websites should, therefore, be exemplary in terms of this requirement in the context of heuristics adopted both as guidelines for the development of websites, as well as those used for research and evaluation of their usefulness using the heuristic method.

BIP websites are always marked with the appropriate logotype. Although a BIP website is linked to the authorities of a given city, it is a separate website and differs from the website of the city's administration unit. Given that, administrators of the respective types of websites (i.e. BIP and city administration, such as <https://www.wroclaw.pl> and <http://bip.um.wroc.pl>), often cooperate by providing hyperlinks to each other's websites or by distributing content according to its function. In some cases, both websites are placed next to each other, i.e. on the same server, but being two different and separate projects.

## V. RESEARCH METHODOLOGY

### A. Procedure

In order to identify heuristics relevant for the assessment of the usefulness of websites of public administration units, a study was conducted to assess the usefulness of selected 60 websites of the Public Information Bulletin. The examination was performed according to the following procedure:

1. Selecting the Public Information Bulletin websites for research purposes.
2. Researching the websites of the Public Information Bulletin using J. Nielsen's heuristics.
3. Identification of basic errors related to the usability of the Public Information Bulletin websites examined.
4. Categorisation of usability errors.

5. Evaluation of the importance and ranking of the different categories of errors.
6. Comparison of the proposed categories of errors with J. Nielsen's heuristics.

The results of the study are presented in the paragraphs below.

*B. Identification of basic errors related to the usability of the Public Information Bulletin websites examined*

The research began with an analysis of sixty websites of the Public Information Bulletin of large Polish cities. The research was conducted by 120 students aged 23-40, during classes in the subject of "usability of the human-computer interface". These people have been trained to do this task. Students worked in two-person groups, each researched one website. The test results were then verified by an expert. The website evaluation procedure was based on Nielsen's heuristics. After determining the general state of the usefulness of websites of this type, the research was narrowed down to twenty largest cities in terms of population (the most up-to-date data from the Central Statistical Office, i.e. from 31.12.2015 were used [32]). This time the analysis was more in-depth due to the fact that it included accessibility aspects. A number of errors and violations were thus identified, which had a material impact on the usability assessment. At a later stage, those had to be classified. A detailed analysis allowed us to identify the areas of the most frequently occurring errors and problems. Fourteen categories of errors were formulated:

**F01. Website ergonomics:** non-intuitive and unusual location of the website's key elements (e.g. main menu, search fields, accessibility functions, etc.) and too large and unstructured accumulation of elements on the main page, including many unnecessary ones.

**F02. Website consistency:** the selective appearance of key elements that should appear on each page within the website (e.g. main menu, footer, search field, etc.).

**F03. Content and its form:** errors in the text (spelling, punctuation, etc.), incorrect encoding of diacritical marks, illegible and inconsistent formatting and arrangement of the text (typefaces, colours, boldening, indentations, spaces, etc.), too few or too many graphic elements (including photographs) affecting the quality of the visitor's website experience, non-standard or user-unfriendly content presentation, and frequent replacement of content with external attachments (e.g. as PDF files).

**F04. Substantive content:** outdated or incomplete information, inconsistency of the information presented within pages belonging to a single category (e.g. selective contact details for individual departments of the city council - telephone and fax numbers provided for some of them and only an e-mail address provided for others), use of a specialist (legal or technical vocabulary) or convoluted (multiply compound sentences, etc.) language.

**F05. Navigation, menus, and grouping of web pages:** too many or too few options in the main or auxiliary menu (the

problem of a proper number of nests), non-intuitive arrangement and illegible presentation of options, lack of clear information about the possibility of rolling down submenus, recurring menu panels across one page, inconsistencies of individual instances of the website's main or auxiliary menu.

**F06. Navigation between web pages:** lack of or errors in breadcrumb navigation, poorly visible navigation panel, inconsistently performing links, lack of return to parent location button, lack of redirection to the homepage after pressing the logotype or title.

**F07. Navigation - website search engine:** performance errors, lack of results, unconventional format of results (e.g. official documents only), lack or a small number of advanced search options (filtering), lack of hints when entering text.

**F08. Navigation - links:** incorrectly described (alternative text) and outdated hyperlinks, references to non-existent locations, lack of description of error 404, lack of information about redirecting to an external website, lack of options for opening new pages in a new tab or in a new window.

**F09. Accessibility - mobile devices:** lack of website responsiveness, incorrectly executed mobile version of the website, problems with scaling individual elements (e.g. search fields).

**F10. Accessibility - colour set:** aesthetically unpleasant shades of colours and their saturation, too big or too small variety of colours, too big or too small contrasts.

**F11. Accessibility - functions:** illegible text, incorrect performance or lack of buttons related to accessibility (e.g. text scaling, changing contrast, etc.).

**F12. Accessibility - website map:** lack of or incorrectly designed, illegible website map.

**F13. Help:** hardly exhaustive or even non-existent help section, errors in the help section (problems which also concern the frequently asked questions), lack of hints and messages in problematic areas of the website.

**F14. Other errors and limitations:** the website loading time is too long or the loading process is completely stopped – often without any messages, access to all functionalities of the website is possible only after registration.

By means of expert analysis, each category of errors was rated in terms of its importance for the overall evaluated of the service's usefulness. The highest ranks were assigned to the categories that determine the possibility of using the website's functionalities, while the lowest ones reflect problems causing only users' moderate discomfort. The scale of the ranks is as follows:

- 1 – a problem of least significance;
- 2 – a minor problem;
- 3 – a problem of average significance;
- 4 – a major error;
- 5 – a critical error.

Each of the identified error areas was assigned one of five ranks. The results of this study are presented in table 1.

TABLE 1.  
RANKS OF ERROR CATEGORIES

No.	Error categories	Rank
F01	Website ergonomics	2
F02	Consistency across the website	2
F03	Content and the form of content presentation	3
F04	Content and the substantive matter	3
F05	Navigation, menu, and page grouping	5
F06	Navigation between web pages	5
F07	Navigation – website search engine	4
F08	Navigation – links	5
F09	Accessibility – mobile devices	4
F10	Accessibility – colour set	3
F11	Accessibility – functions	5
F12	Accessibility – website map	1
F13	Help	1
F14	Other error and hindrances	3

The most serious problems (rank 5 and 4) found across the BIP websites under examination are navigation difficulties (F05-F08) and availability limitations (F09 and F11). Violations such as F05-F08 i.e. ones related to website navigation can make it completely impossible to find the information needed by the user. During testing, in many cases, the unintuitive menu layout, containing an enormous number of mixed and unnecessary options, combined with an unoperational search engine, made it impossible to find the searched content.

The second type of serious error concerns availability. An increasing number of people are using smartphones and tablets, more and more often abandoning desktop computers. The lack of possibility to use a mobile device or limitations in this respect may effectively discourage many Internet users. Also the lack of accessibility-related functions (e.g. change of contrast) means a serious barrier for people with medical conditions, thus striking the basic principles and sense of BIP websites. The importance of colour choices (F10) has been rated as slightly lesser (average rank, i.e. 3) as it is solved by the contrast matching option mentioned above. Moreover, in none of the cases analysed did the colour scheme pose a considerable problem when reading the content. The same rank was assigned to F03 and F04. These are important aspects of a website, but rather than preventing its use they result in the user's impatience and irritation. The last area, F14, was also given an average rating, due to the diversity and occasionality of errors. The first two categories are less important for the perception of the website and are associated with bad user experience rather than serious impairment of usability, therefore they were assigned a lower rank of 2. The least important are areas F12 and F13, which should be only a supplement to a well-developed website.

### C. Comparison of the identified categories of errors with J. Nielsen's heuristics

The errors identified across the BIP websites, described in the previous section (i.e. F01-F14), were assigned to Jakob Nielsen's ten heuristics (specified in section 3). Table 2 presents a breakdown of the identified error areas and Nielsen's heuristics.

Every link between an error category and a heuristic is marked with an "X". One error area can be associated with several heuristics, while one heuristic can cover several of the specified problem categories. The last column and last row of the table summarise the number of links. For example, category F05 is thematically linked with almost all heuristics (from H01 to H08). Hence, the sum at the end of the line (last column) is 8. This category, therefore, affects many aspects of the site. The table can also be read from the perspective of heuristics. For example, H09 will only be affected in four error categories (F08, F11, F13, F14).

H03-H05 and H07 are the most frequently violated heuristics, i.e. those linked with the highest number of errors. These relate to the user's control over navigation, maintaining consistency and standards, error prevention, as well as errors that affect the effectiveness of use. These are therefore heuristics concerning the most serious usability violations. As for the error areas that concern the greatest number of heuristics, these are as follows: F05, F08, and F11, i.e. again navigation and availability and the less important element of B13 (help).

It turns out that the most serious errors are also the most common ones: F05-F08 (navigation, menu, control, search engine, and links) and F11 (accessibility functions). Individual usability violations covered by these areas appeared in at least half of the websites examined.

## VI. CONCLUSIONS AND FUTURE WORKS

This paper presents identified categories of errors occurring on public administration websites, which were associated with traditional heuristics of Jakob Nielsen. The proposed categories may be heuristics to be applied under the heuristic method for assessing the usefulness of websites. Compared to J. Nielsen's heuristics, in the identified categories of errors, many refer to website accessibility. Meeting the requirements for website accessibility is ensured by features such as the clarity and intuitiveness of the website, which translates into a good reception of the website by both healthy and disabled people. This means that the greater accessibility of the website improves usability as perceived by all users. Therefore, when testing the usability of a website, it is necessary to pay more attention to the verification of its availability.

Identification of the most frequent errors and usability violations on the websites of public administration units as well as determining their correlation with Jakob Nielsen's heuristics will be the basis for further research in this area. Its aim is to develop a comprehensive procedure for testing the usability and availability of public administration

services, with expert analysis being an element of key importance. Tests using this procedure will be conducted on the websites of the Public Information Bulletin. Further work may also result in the presentation of a modified

version of Nielsen's classic heuristics, tailored to the needs of testing public websites. The research is needed especially in view of those at risk of digital exclusion, as well as in view of the rapid growth of the Internet.

TABLE 2.  
HEURISTICS VIOLATIONS

Error categories	No.	Jakob Nielsen's ten heuristics										Number of heuristics linked to the error	
		Visibility of system status H01	Correspondence between the system and the real world H02	User control and freedom H03	Consistency and standards H04	Prevention of errors H05	Recognition and not remembering H06	Flexibility and efficiency of use H07	Aesthetic and minimalist design H08	Helping users recognise and diagnose errors, and enabling recovery from errors H09	Help and documentation H10		
Website ergonomics	F01	X		X	X			X	X				5
Consistency across the website	F02	X		X	X	X		X					5
Content and the form of content presentation	F03		X		X	X			X				4
Content and the substantive matter	F04		X		X	X	X					X	5
Navigation, menu, and page grouping	F05	X	X	X	X	X	X	X	X				8
Navigation between web pages	F06	X		X	X	X	X	X					6
Navigation – website search engine	F07		X	X	X	X		X				X	6
Navigation – links	F08	X		X	X	X		X		X		X	7
Accessibility – mobile devices	F09			X	X	X		X	X				5
Accessibility – colour set	F10				X				X				2
Accessibility – functions	F11		X	X	X	X		X	X	X		X	8
Accessibility – website map	F12	X		X	X		X	X	X				6
Support	F13	X	X	X		X	X	X		X		X	8
Other error and hindrances	F14	X		X		X		X		X			5
Number of heuristic violations													
		8	6	11	12	11	5	11	7	4	5		

REFERENCES

[1] ISO 9241-210: 2010. Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems. <https://www.iso.org/obp/ui/#iso:std:iso:9241:-210:ed-1:vl:en:en,%20dostep%20dnia%2013.05.2016> (12.01.2018).

[2] Quiñones, D., Rusu, C., 2017. How to develop usability heuristics: A systematic literature review. *Computer Standards & Interfaces*, vol. 53, pp. 89–122. doi: 10.1016/j.csi.2017.03.009.

[3] Inostroza, R., Rusu, C., Roncagliolo, S., Rusu, V., Collazos, C. A., 2016. Developing SMASH: A set of smartphone’s usability heuristics. *Computer Standards & Interfaces*, 43, pp. 40-52.

[4] Nielsen, J., 2012. Usability 101: Introduction to Usability. <https://www.nngroup.com/articles/usability-101-introduction-to-usability> (29.10.2018).

[5] Khalid H., Hedge A., Ahram T., *Advances in Ergonomics Modeling and Usability Evaluation*, CRC Press 2011.

[6] Paplauskaitė, L., 2014. Usability and usefulness in UX Web Design. <https://bitzesty.com/2014/05/15/usability-and-usefulness-in-ux-web-design> (20.07.2018).

[7] Lawton Henry, S., 2019. Introduction to Web Accessibility. <https://www.w3.org/WAI/fundamentals/accessibility-intro> (1.05.2019).

[8] Robbins, J. N., 2014. Projektowanie stron internetowych, Przewodnik dla początkujących webmasterów po HTML5, CSS3 i grafice. Wydawnictwo Helion, Gliwice.

[9] Phyo, A., 2003. Web Design, Projektowanie atrakcyjnych stron WWW. Wydawnictwo Helion, Gliwice.

[10] Streich, S., Accessibility is NOT just for people with disabilities. <https://vimm.com/website-accessibility> (1.05.2019).

[11] Morville, P., 2004. User Experience Design. [http://semanticstudios.com/user\\_experience\\_design](http://semanticstudios.com/user_experience_design) (20.07.2018).

[12] Waddell, C., Regan, B., Lawton Henry, S., Burks, M. R., Thatcher, J., Urban, M. D., Bohman, P., 2002. *Constructing Accessible Web Sites*. Apress Publishing House.

[13] Lawton Henry, S., Abou-Zahra, S., White, K., 2016. Accessibility, Usability, and Inclusion. <https://www.w3.org/WAI/fundamentals/accessibility-usability-inclusion> (1.05.2019).

[14] Fernandez, A., Insfran, E., Abrahão, S., 2011. Usability evaluation methods for the web: A systematic mapping study, *Information and Software Technology*, vol. 53, Issue 8, pp. 789-817. doi.org/10.1016/j.infsof.2011.02.007.

[15] Lazar, J., Feng, J. H., Hochheiser, H., 2010. *Research Methods in Human-computer Interaction*. John Wiley & Sons.

[16] Paz, F., Pow-Sang, J. A., 2016. A Systematic Mapping Review of Usability Evaluation Methods for Software Development Process. *International Journal of Software Engineering and Its Applications*, Vol. 10, No. 1 (2016), pp. 165-178. doi: 10.14257/ijseia.2016.10.1.16.

[17] Sikorski, M., 2012. *User-system Interaction Design in IT Projects*. Gdańsk University of Technology.

[18] Tullis, T., Albert, B., 2008. *Measuring the User Experience. Collecting, Analyzing, and Presenting Usability Metrics*. Morgan Kaufmann Publishers, Amsterdam.

[19] Scholtz, J., 2004. Usability evaluation. [http://notification.etisalat.com/eg/etisalat/templates.backup.16082011/582/Usability%2520Evaluation\\_rev1%5B1%5D.pdf](http://notification.etisalat.com/eg/etisalat/templates.backup.16082011/582/Usability%2520Evaluation_rev1%5B1%5D.pdf).

[20] Philips, M., 2017. Elevate Your UX with a Heuristic Analysis – How to Run a Usability Evaluation. <https://www.linkedin.com/pulse/elevate-your-ux-heuristic-analysis-how-run-miklos-philips> (15.09.2018).

- [21] Gerhardt-Powals, J., 1996. Cognitive engineering principles for enhancing human-computer performance. *International Journal of Human-Computer Interaction*. Volume 8 Issue 2, April-June 1996, pp. 189-211.
- [22] Weinschenk, S., Barker, D. T., 2000. *Designing Effective Speech Interfaces*. Wiley.
- [23] Shneiderman, B., 2006. *The Eight Golden Rules of Interface Design*. *Designing the User Interface*. 6th Edition, Section 3.3.4.
- [24] Inostroza, R., Rusu, C., Roncagliolo, S., Jimenez, C., Rusu, V., 2012. Usability Heuristics for Touchscreen-based Mobile Devices. *IEEE Xplore*, DOI: 10.1109/ITNG.2012.134.
- [25] Tognazzini, B., 2014. *First Principles of Interaction Design (Revised & Expanded)*. askTog. <https://asktog.com/atc/principles-of-interaction-design> (9.05.2019).
- [26] Douglas, S., 2017. 7 Usability Heuristics That All UI Designers Should Know. *Usability Geek*. <https://usabilitygeek.com/usability-heuristics-ui-designers-know> (19.01.2019).
- [27] Jimenez, C., Lozada, P., Rosas, P., 2016. Usability heuristics: A systematic review. In: 11th Colombian Computing Conference, 27-30.09.2016, Popayan, Colombia, doi: 10.1109/ColumbianCC.2016.7750805.
- [28] Dourado, M. A. D., Canedo E. D., 2018. Usability Heuristics for Mobile Applications – A Systematic Review. In: *Proceedings of the 20th International Conference on Enterprise Information Systems (ICEIS'2018)*, vol. 2, pp. 483-494 doi:10.5220/0006781404830494.
- [29] Section 508 of the Rehabilitation Act. - 29 U.S.C. § 798. Section 508 - Electronic and Information Technology. <https://www.fcc.gov/general/section-508-rehabilitation-act> (20.02.2019)
- [30] Barrierefreie Informationstechnik-Verordnung — BITV 2.0. <https://www.barrierefreies-webdesign.de/bitv/bitv-2.0.html> (20.02.2019)
- [31] Accessibilità siti web. AgID promuove l'accessibilità dei siti web in relazione alla normativa vigente. <https://www.agid.gov.it/it/design-servizi/accessibilita-siti-web> (20.02.2019)
- [32] Miasta największe pod względem liczby ludności. Główny Urząd Statystyczny. <https://stat.gov.pl/statystyka-regionalna/rankingi-statystyczne/miasta-najwieksze-pod-wzgle-dem-liczby-ludnosci> (1.02.2019)



# Motivations for BPM Adoption: Initial Taxonomy based on Online Success Stories

Renata Gabryelczyk  
University of Warsaw  
Faculty of Economic Sciences  
ul. Długa 44/50,  
00-241 Warszawa, Poland

Email: r.gabryelczyk@wne.uw.edu.pl

Aneta Biernikowicz  
University of Warsaw  
Faculty of Management  
ul. Szturmowa 1/3  
02-678 Warszawa, Poland

Email: abiernikowicz@wz.uw.edu.pl

□

**Abstract** — The main aim of this research in progress is to develop an initial taxonomy of motivations underlying BPM (Business Process Management) adoption in organizations. This initial study is based on the analysis of 75 customer cases and success stories published on-line by BPM system vendors and BPM consulting companies. We used the mixed conceptual/empirical approach to taxonomy development basing the empirical analysis on descriptive data-coding canon. As the result of our research we present an initial taxonomy of the motivations for the adoption and use of BPM that consists of three dimensions: the organizational scope of a BPM initiative (enterprise-wide, process-focused or task oriented); presence (or not) of the information technology component; and, the importance of external versus internal drivers motivating a BPM initiative. Proposed initial taxonomy will be developed in further research and will serve to link the motivation to change with the expected benefits of BPM adoption.

## I. INTRODUCTION

Any organization making a decision about the adoption of Business Process Management (BPM) is guided by their specific motivations. These motivations are understood as the main reasons why organizations take BPM initiatives as a set of arguments used to support a decision concerning BPM implementation. The motivations behind the justification of organizational change are an indispensable element of every business case [1]. It is these motivations that most often reflect the benefits of BPM adoption [2]. In this study, the term “*motivation*” is related to its goals and expected benefits being the starting point for the decision to adopt BPM.

Although research generally confirms that individual motivations of employees are translated into the performance of the entire organization [3-5], studies on motivations for BPM adoption are virtually non-existent.

For the purpose of better understanding, analysis and use in future studies, motivations should be organized and classified by groups. In this study, we plan to develop an initial taxonomy of motivations which, according to the literature, can serve as a form of classification and as “*a fundamental mechanism for organizing knowledge*” [6, pp.

11-12]. Taxonomies help to arrange concepts, to perceive the relations between concepts and to draw conclusions from them. Taxonomies also reduce complexity, which is why they are useful and important for both research and management practice [6, 7].

In our study, we will use the methodology for taxonomy development proposed by Nickerson et al. [6] that is established in the field of Information Systems. However, it will be the first time applied in BPM research.

To develop an initial taxonomy of BPM motivations, we will use secondary data from BPM case studies and success stories published on-line. We believe that the identification and initial taxonomy of BPM motivational factors will bring a new and original contribution not only to BPM research but also to the practice of the planning of BPM adoptions.

This paper will be organized as follows: firstly, the research background will be presented, including theories underlying BPM as well as short discussions from the literature on the benefits of BPM. This theoretical background will be followed by the explanation of the research methodology used to create a taxonomy. The research results obtained will then be presented specifically as an initial taxonomy. Finally, the contribution and limitations of the study will be assessed and the direction of future research proposed.

## II. RESEARCH BACKGROUND

### *Theories underlying BPM motivations*

Theories and frameworks used for explaining BPM can help identify potential motivations and related goals to achieve through the use of BPM. Starting from the roots of process-based management concepts, we can point out two main perspectives in looking at BPM and the expected outcomes of its adoption: *the organizational perspective* and *the technological perspective*. For the organizational perspective, research and practice were focused on using process thinking during the design and improvement of an organization [8-10]. The technological perspective was

□ This work was supported by the Polish National Science Centre, Poland, Grant No. 2017/27/B/HS4/01734

addressed by using process-based concepts and tools to support the design and implementation of IT systems [11, 12].

An integrated and interdisciplinary BPM framework was proposed by [13, 14], who indicated *six core BPM elements* required for the holistic and sustainable use of process management. These include strategic alignment, governance, methods, information technology, people and culture. Further studies on BPM also began to emphasize the importance of the contextual and environmental factors for BPM adoption [15, 16]. As the BPM concept became more established, the list of potential expected benefits of BPM adoption expanded. Motivational factors could also be driven by customers, suppliers, competitors and legal pressures exerted on organizations [16, 17].

To explain BPM phenomena in an organization, the BPM literature indicates mostly theories, which we also present in this study as the main theories underlying BPM adoption and enhancing understanding of BPM motivations [18, 19].

The *technology-task fit theory* is mainly used in the field of Information Systems and explains the relationship between processes and technology. According to this theory, benefits from the implementation and use of IT systems in organizations can be gained if the information system fits the tasks that need to be performed [20]. This theory can explain motivational factors related to the use of technology. The *dynamic capabilities theory* refers to the purposeful adaptation of organizational resources and competencies in the continual improvement process to respond better to a changing environment [21]. The *contingency theory* points to the situational fit between the method of organizing and managing and the environment in which the organization operates [22, pp. 96-100]. Thus, the contingency theory explains the aspect of environmental motivations which are forced by the external environment.

To enhance the understanding of BPM phenomena the theories referred to above should be synthesized [19, 23]. This integrated approach can serve as a common platform for developing a comprehensive theory explaining BPM.

#### *Motivations as benefits drivers*

The adoption of any new approach or organizational change like BPM is an effort for an organization. This effort must be justified by the expected benefits that should result from the investment of its effort. Therefore, studying factors that motivate organizational change, including BPM adoption, should consider the analysis of the perceived organizational benefits of the implemented change [2, 24, 25].

We propose to discuss the connection between the set of motivations and goals and the set of outcomes and benefits. We understand the term "*benefit*" as the desirable and measurable outcome of BPM implementation, where outcomes are "*the goals a company realized*". We understand the goals as "*something a company desires to achieve*" [24], whereas motivations are primary reasons that inspire an organization to adopt BPM. The connecting elements between the two aforementioned sets are planned activities.

We can formulate the following chain of connections that explain why an organization's motivations should be studied: Motivations → Goals → Activities → Outcomes → Benefits

The study of Malinova et al. [24] explains an important relationship between the goals articulated for a BPM initiative and its actual outcomes and benefits. The vehicle that delivers outcomes inspired by the goals is a set of activities that a company undertakes within the scope of a BPM initiative.

The expectation of benefits may encourage decision makers to give support to a BPM initiative in their organization and would shape their expectations as to what can be achieved with it. On the basis of these expectations, combined with their assessment of current organizational needs, the goals of a BPM initiative are formulated. Depending on the general goal of a BPM initiative a different approach to its realization could be taken - a more centralized, top-down approach focused on managing few processes at the time, or, a decentralized approach with multiple distributed initiatives relying more on a dynamic organizational social system. We can, therefore, conclude that knowledge of the initial motivations of BPM will contribute to the success of the undertaking initiative.

### III. METHODOLOGY

#### *On-line cases collection and coding*

In order to collect data for our analysis we searched for published on-line BPM cases and success stories using the following search strings:

- Search string I ("Business Process Management" OR "BPM" AND "case study")
- Search string II ("Business Process Management" OR "BPM" AND "success story")

The collected cases were used as secondary data [26]. As this study is preliminary, we limited the number of stories by choosing recurring websites in both search strings and diversified them by choosing three websites of BPM suites vendors and three websites of BPM consulting companies. In total, 75 BPM case studies were used to propose an initial taxonomy. Due to the fact that proposed taxonomy can be further developed and we do not present the results of quantitative research, we believe that this sample is sufficient to present the study in progress.

In the first step of the analysis, in each case, we identified excerpts that offered reasons why the organization decided to adopt BPM. We identified 271 individual items of motivation which were subsequently coded based on descriptive data-coding canon [27]. We used NVivo software to support the coding process and analysis of qualitative data.

#### *Taxonomy development process*

We applied the methodology for taxonomy development by Nickerson et al. [6] and according to this study, we determined the meta-characteristic of motivations as the most comprehensive, based on theories underlying BPM adoption. We used three main characteristics as the basis: motivation

driven by an organization, by technology, and by the environment. The further development of taxonomy, therefore, be based on this conceptual pillar.

Due to the current lack of useful classifications of BPM motivation in the literature, we elected to use the mixed conceptual/empirical approach to taxonomy development [6]. We employed an empirical approach using coding canon to row data i.e., descriptions of motivations identified in each case. Subsequently, we coded and classified similar data under the same category using a deductive method of conceptualizing of data.

We made every effort to meet the qualitative conditions for taxonomy, which should be concise, robust, comprehensive, extendible, and explanatory [6]. However, the number of dimensions and the number of motivations can be extended in future research, so our preliminary taxonomy may be less than comprehensive.

#### IV. RESULTS

According to the used methodology [6], we identified three main dimensions to develop this initial taxonomy of BPM motivations. Based on BPM knowledge, we proposed characteristics for each dimension.

Analyzing the BPM literature on objectives, outcomes and the overall benefits of BPM, we noticed that they are formulated at different organizational levels. At the 'macro' level of the organization they refer to the overall effectiveness, strategy, organizational structures, methods of allocation and utilization of resources, etc. At the level of

processes, the goals and outcomes often refer to one or few more processes or the phases of the process life cycle, such as process design, analysis, redesign, implementation, monitoring and controlling [24]. Occasionally, the formulation of goals and expected benefits is focused on even more specific elements such as work positions and relate to tasks. A good example of such a situation may be the formulation of goals for the recently popular Robotic Process Automation applications where outcomes are expected at task or process levels [28]. Although it is obvious that all achieved results translate into the effectiveness of the entire organization [9], focusing the motivation either on the entire organization or process, or task will indicate the scope and complexity of planned organizational changes. For this reason, we decided to highlight the scope characteristics and three levels of impact within the 'Organization' dimension.

When planning the second dimension, 'Technology', we took into account the long-term relationship of the applied process approach with the implementation and use of information technology, which was also highlighted by us in the background of this study. We have, therefore, decided to examine to what extent the motivations for adopting BPM are inspired by technology.

In the third dimension, 'Environment', our intention was to check whether it is the internal or external environment that motivates decisions regarding BPM adoption more often. The internal environment is shaped mainly by the organization's owners, the board of directors, employees and organizational culture. However, BPM initiatives may also be triggered by

TABLE I.  
INITIAL TAXONOMY OF BPM MOTIVATIONS

<i>Examples of coded motivations</i>	Motivations related to the scope of the BPM initiative in an Organization			Motivations related to the use of Technology		Motivations driven by Environment	
	Motivations driven at the organizations level	Motivations driven at the process level	Motivations driven at the task level	Techno-logical motivations	Non techno-logical motivations	Internal environment driven motivations	External environment driven motivations
Clarify roles and responsibilities	x				x	x	
Improve governance mechanisms	x				x	x	
Comply with new regulatory requirements	x				x		x
Capture organizational knowledge	x				x	x	
Respond to customer requirements	x				x		x
Improve collaboration in an organization	x				x	x	
Eliminate process errors		x			x	x	
Reduce process costs		x			x	x	
Improve financial performance	x				x	x	
Improve process efficiency		x			x	x	
Reduce manual tasks			x	x		x	
Automate tasks			x	x		x	
Implement new IT system	x			x		x	
Improve data and information quality		x			x	x	

pressure from the external environment such as customers, suppliers or changes in regulations [23].

Table I presents the initial taxonomy of BPM adoption. In the first column we present the examples of encoded motivations from our study in order to show that, following the used methodology, at least one motivation is classified under every characteristic of every dimension. Such a taxonomy, according to qualitative attributes, is extensible and explanatory [6]. Therefore, can be extended in future studies.

#### V. CONCLUDING REMARKS, LIMITATIONS AND FUTURE RESEARCH

It seems obvious that organizations are motivated to adopt BPM by the desire to achieve expected benefits. However, the primary motivations for this adoption remained unexplored until the results of this study were presented. Our research is the first attempt to identify and categorize the initial reasons that grounded the decision on the adoption of BPM. Our research contributes to the theory by offering the first methodologically developed taxonomy of BPM motivation.

We believe that our research will contribute to inspiring organizations and dispelling their possible doubts about the benefits of adopting BPM.

We also acknowledge some limitations of our study. Firstly, it might be the case that success stories published on vendors' websites tend to describe cases with only positive effects of BPM application. Secondly, the investigated cases described the effects of various process initiatives without distinguishing between one-time improvement projects and initiatives that are a part of a systematic approach to BPM adoption that could enable ongoing process-based management of an organization. Motivations for these types of projects may have different dynamics.

In future studies, we plan to extend the sample of success stories. Based on classified motivations, we intend to develop motivation patterns that will include groups of motivating factors and organization characteristics. Finally, we plan to investigate how the various types of motivations impact the future success or failure of BPM initiatives.

#### REFERENCES

- [1] Rudden, J.: Making the case for BPM: a benefits checklist. *BPTrends* (2007)
- [2] Malinova, M., Mendling, J.: A qualitative research perspective on BPM adoption and the pitfalls of business process modeling. In: La Rossa, M., Soffer, P. (eds.) *Business Process Management Workshops*, vol. LNBP 132, pp. 77-88. Springer, Berlin-Heidelberg (2013)
- [3] Moon, M.J.: Organizational Commitment Revisited in New Public Management: Motivation, Organizational Culture, Sector, and Managerial Level. *Public Performance & Management Review* 24, 177-194 (2000). <https://doi.org/10.2307/3381267>
- [4] Paarlberg, L.E., Lavigna, B.: Transformational leadership and public service motivation: Driving individual and organizational performance. *Public Administration Review* 70, 710-718 (2010) <https://doi.org/10.1111/j.1540-6210.2010.02199.x>
- [5] Haque, M.F., Haque, M.A., Islam, M.: Motivational Theories-A Critical Analysis. *ASA University Review* 8, 61-68 (2014)
- [6] Nickerson, R.C., Varshney, U., Muntermann, J.: A method for taxonomy development and its application in information systems. *European Journal of Information Systems* 22, 336-359 (2013) <https://doi.org/10.1057/ejis.2012.26>
- [7] Bailey, K.D.: *Typologies and taxonomies. An Introduction to classification Techniques*. Sage, Thousand Oaks CA (1994)
- [8] Hammer, M., Champy, J.: *Reengineering the Corporation. A Manifesto for Business Revolution*. Harper Business (1993)
- [9] Rummler, G.A., Brache, A.P.: *Improving performance: how to manage the white space on the organization chart*. Jossey-Bass (1995)
- [10] McCormack, K.P., Johnson, W.C.: *Business Process Orientation. Gaining the E-Business Competitive Advantage*. CRC Press (2001)
- [11] Scheer, A.-W., Jost, W.E.: *ARIS in der Praxis. Gestaltung, Implementierung und Optimierung von Geschäftsprozessen*. Springer, Berlin, Heidelberg (2002)
- [12] Van der Aalst, W.M.P., Ter Hofstede, A.H.M., Weske, M.: *Business Process Management: A Survey*. In: van der Aalst, W.M.P., Weske, M. (eds.) *Business Process Management*, vol. LNCS 2678, pp. 1-12. Springer, Berlin, Heidelberg (2003)
- [13] De Bruin, T., Rosemann, M.: Towards a Business Process Management Maturity Model. In: Bartmann, D., Rajola, F., Kallinikos, J., Avison, D., Winter, R., Ein-Dor, P., al., e. (eds.) *Proceedings of the Thirteenth European Conference on Information Systems*, 26-28 May 2005, Germany, Regensburg (2005)
- [14] Rosemann, M., Vom Brocke, J.: The six core elements of Business Process Management. In: vom Brocke, J., Rosemann, M. (eds.) *Handbook on Business Process Management 1*, pp. 105-122. Springer (2015)
- [15] Vom Brocke, J., Zelt, S., Schmiedel, T.: On the role of context in Business Process Management. *International Journal of Information Management* 36, 486-495 (2016) <https://doi.org/10.1016/j.ijinfomgt.2015.10.002>
- [16] Gabryelczyk, R., Roztocki, N.: Business Process Management success framework for transition economies. *Information Systems Management* 35, 234-253 (2018) <https://doi.org/10.1080/10580530.2018.1477299>
- [17] Willaert, P., Van Den Bergh, J., Willems, J., Deschoolmeester, D.: The process-oriented organisation: a holistic view developing a framework for business process orientation maturity. In: G., A., P., D., M., R. (eds.) *Business Process Management*, vol. LNCS 4714, pp. 1-15. Springer (2007)
- [18] Van Looy, A., Van den Bergh, J.: The Effect of Organization Size and Sector on Adopting Business Process Management. *Business & Information Systems Engineering* 60, 479-491 (2018)
- [19] Trkman, P.: The Critical Success Factors of Business Process Management. *International Journal of Information Management* 30, 125-134 (2010) <https://doi.org/10.1016/j.ijinfomgt.2009.07.003>
- [20] Furneaux, B.: Task-Technology Fit Theory: A Survey and Synopsis of the Literature. In: Dwivedi, Y.K., Wade, M.R., Schneberger, S.L. (eds.) *Information Systems Theory: Explaining and Predicting Our Digital Society*, Vol. 1, pp. 87-106. Springer New York, New York, NY (2012)
- [21] Teece, D.J., Pisano, G., Shuen, A.: Dynamic capabilities and strategic management. *Strategic Management Journal* 18, 509-533 (1997)
- [22] Scott, R.W.: *Organizations. Rational, Natural, and Open Systems*. Prentice Hall, New Jersey (2003)
- [23] Gabryelczyk, R.: Exploring BPM Adoption Factors: Insights into Literature and Experts Knowledge. In: Ziemba, E. (ed.) *Information Technology for Management: Emerging Research and Applications*, vol. LNBP 346, pp. 155-175. Springer International Publishing, Cham (2019)
- [24] Malinova, M., Hribar, B., Mendling, J.: A framework for assessing BPM success. *Proceedings of the 22nd European Conference on Information Systems*. Association for Information Systems, Tel Aviv, Israel (2014)
- [25] Tuček, D.: The Main Reasons for Implementing BPM in Czech Companies. *Journal of Competitiveness* 7, 126-142 (2015)
- [26] Poba-Nzaou, P., Uwizeyemungu, S., Raymond, L., Paré, G.: Motivations underlying the adoption of ERP systems in healthcare organizations: Insights from online stories. *Information Systems Frontiers* 16, 591-605 (2012)
- [27] Saldana, J.: *The Coding Manual for Qualitative Researchers*. Sage, Los Angeles, London (2013)
- [28] Houy, C., Hamberg, M., Fettke, P.: Robotic Process Automation in Public Administrations. In: Räckers, M., Halsbenning, S., Rätz, D., Richter, D., Schweighofer, E. (eds.) *Digitalisierung von Staat und Verwaltung*, pp. 62-74. Gesellschaft für Informatik eV, Bonn (2019)

# Multi-criteria approach to viral marketing campaign planning in social networks, based on real networks, network samples and synthetic networks

Artur Karczmarczyk\*, Jarosław Jankowski\* and Jarosław Wątróbski†

\*Faculty of Computer Science and Information Technology

West Pomeranian University of Technology in Szczecin, Żołnierska 49, 71-210 Szczecin, Poland

Email: {artur.karczmarczyk,jaroslaw.jankowski}@zut.edu.pl

†Faculty of Economics and Management

University of Szczecin

Mickiewicza 64, 71-101 Szczecin, Poland

Email: jwatrobski@usz.edu.pl

**Abstract**—Spreading of information within social media and techniques related to viral marketing take more and more attention from companies focused on targeting audiences within electronic systems. Recent years resulted in extensive research centered around spreading models, selection of initial nodes within networks and identification of campaign characteristics affecting the assumed goals. While social networks are usually based on complex structures and high number of users, the ability to perform detailed analysis of mechanics behind the spreading processes is very limited. The presented study shows an approach for selection of campaign parameters with the use of network samples and theoretical models. Instead of processing simulations on large network, smaller samples and theoretical networks are used. Results showed that knowledge derived from relatively smaller structures is helpful for initialization of spreading processes within the target network of larger size. Apart from agent based modeling, multi-criteria methods were used for evaluation of results from the perspective of costs and performance.

## I. INTRODUCTION

Online platforms evolved from early stage technical systems to social media with integrated mechanics of social communication and interactions close to the real world [1]. Together with growing audiences, they attracted more attention of marketers. Apart from typical digital marketing channels based on display advertising and search engines new strategies focused on social media emerged. They include mechanism based on detailed targeting, consumer behavior analysis and commercial content dissemination with the use mechanisms of information spreading.

Results delivered from viral campaigns usually outperform traditional campaigns because of the utilized social influence and ability to induce high dynamics even with low budgets [2]. Social recommendations have high impact on customer decisions and, properly integrated with marketing communication [3], help to further increase performance [4].

The recent studies focused on viral marketing take into account data from real platforms as well as theoretical network models [5]. One of the goals is to increase campaign dynamics

and coverage with properly selected initial customers during the seeding process [6]. Apart from static networks, dynamic networks with varying structures are taken into account [7]. Other approaches take into account multi-layer structure of networks representing specifics of real social relations based on different networks, for example private and professional contacts [8].

Theoretical and simulation models are used for prediction of network coverage. They can be derived from analytic models used in epidemiology [9] or can be more focused on network structures and characteristics [10]. Other possibility is to use theories and models related to the diffusion of innovations [11].

While most of the research is focused on coverage and number of infected nodes within the network, from the practical point of view, marketing campaigns can have different goals and specifics. They are planned within assumed budget constraints and timing. A different strategy can be used to acquire high number of potential customers in a very short time than for a long term planning and organic growth of customer database. Campaign budget influences the number of initially infected nodes (seeds) and demographic characteristics. The quality of seeds and their number can be a key factor of campaign coverage and overall results. Additional budgets can be used to increase campaign dynamics or lifespan. To take into account various goals multi-criteria campaign evaluation can be used to select campaign parameters and goals according to preferences and priorities [12]. Earlier research has shown that in order to reduce computational complexity, campaigns can be planned with the use of simulations within smaller synthetic networks based on theoretical model. However, since the theoretical models might not always fit the real networks, the current study proposes the use of network samples for the initial simulations and detection of campaign parameters. Both approaches were compared with results obtained from the complete network and showed the ability to obtain approximate results with network samples.

The paper comprises of five main sections. After this introduction, in Section II literature review is presented. It is followed by the methodology presented within Section III and results in Section IV. Paper is concluded in Section V.

## II. LITERATURE REVIEW

Social platforms gather detailed information about user behavior and social relations with the main goal to better address commercial messages and properly target products and services [13]. The growing complexity and volumes of the collected data is a direct result of the growing number of users and that their activities moved to electronic systems [14], [15]. Social platforms are treated as tools to use social influence mechanisms to spread information between friends with the impact strengthened by social recommendations. Contacts within social networks are used to pass the information and it often induces information cascades as a main driver of viral marketing campaigns. Multidisciplinary nature of phenomena connected with information diffusion integrates efforts from scientists from various fields like sociology, computer science, physics and management with a different theoretical and practical goals [4] [9] [6].

For better understanding of the information spreading processes, theoretical models are used and they are often implemented within agent based environment or used for analytic studies [16]. Methodological background of studies is often based on models initially created for epidemic research like SIR or SIS with taken into account analytic view on processes and their dynamics [9]. Apart from them, more dedicated solutions were created to create models on microscopic level using information about network structures and relations between users. They are based on two key mechanisms represented by linear threshold models [11] and independent cascades [10]. Linear threshold model, with its later extensions, assumes the social influence induced by neighbors with the network and information flow when the number of neighbors exceeds assumed threshold. Cascading models use different mechanics with spreading based on propagation probabilities and communication with surrounding neighbors and passing content to them. These approaches can be treated as pull and push spreading models. Spreading models can be also used for analysis based on aggregated and macroscopic level [17].

Apart from the mechanics of the information spreading, the dynamics of processes are related to network models and their structures. For the simplest approaches, static networks of non-varying structures are used. More closer to reality are approaches focused on dynamic networks with a changing number of social connections or availability of nodes [18]. For better representation of real systems multi-layer networks are used with spreading dependent on connections between layers, their structures or similarities [8].

Many studies related to information spreading take into account the selection of initial customers, in a form of a seeding process, targeted with product samples or other marketing content with the main goal to motivate them to spread the information to friends within the network [6]. Proper

selection of seeds is crucial for successful campaigns, but the problem identified as influence maximization problem is NP-hard [10]. Greedy solutions deliver effective results, but with the high computational cost they are difficult to use within real networks [10]. More practical approaches base on heuristics and a selection of nodes with the use of the network metrics like degree or betweenness. Centrality measures can be used for selection of initial influencers with assumed characteristics [19] [20].

Apart from seeding only once at the beginning of the process, knowledge about the process performance can be gathered and used for additional actions to improve the process characteristics. Adaptive approaches can be used [21] to increase the reach and better utilize the available knowledge. Other possibility is to spread the seeds over the time and better utilize the natural spreading processes. It can be applied in a form of sequential seeding [22] or its extension with recomputed nodes' rankings at every simulation step [23]. Further improvement of seeding can be performed with the use of knowledge about community structures within the network [24], voting mechanics [25] or k-shell based approach dedicated for identification of nodes with high spreading potential [26].

Apart from single campaigns spreading, processes can interact or compete [27]. For such scenarios seeding can be planned to increase the chance of process to survive among competitors or reach audiences in a shortest time before other processes acquire them. Similar situation takes place in epidemic research where two or more pathogens are competing with each other or conditional infections are observed with activity of first virus required for next viruses. Competing scenarios are observed when awareness spreading is decreasing dynamics of epidemic [28]. It lead to extension of the single campaign models to multi-spreading processes for viral marketing studies[29].

Another studies take into account content specifics and network structures [30], proper ways to motivate users to forward the content [31], influence of emotions on content propagation processes [32] [33] and other structural or functional factors [34] [35].

The earlier studies focused mainly on influence maximization to increase coverage within the network. Campaign evaluation was also discussed as a multi-criteria problem [12]. Campaigns performed within agent based simulation environment were evaluated with the use of set of criteria related to budgets, campaign costs and the number of target nodes. Model output was delivering solutions with defined number of seeds or propagation probabilities. Study also showed the ability to perform simulations with theoretical models and apply selected strategies to real network. The current study extends the presented approach and uses network samples created with the use of snowball sampling [36].

## III. METHODOLOGY

Viral marketing campaigns can be based on various strategies. During the campaign planning, decisions are taken about

optimal number of initial seeds, methods used for their selection, motivation techniques used for users to increase their willingness to spread the content and type of incentives used to increase the propagation probabilities. Similar problems are related to campaign evaluation and selection of campaign metrics dependent on campaign goals. Other performance metrics can be used for campaigns focused on high network coverage than on highly targeted processes addressed only to specific customers.

While social networks store information about users, connections and network structures, it is possible to analyze information before campaign to optimize the strategy and maximize results. With the assumed campaign scenarios and goals it is possible to simulate and test different strategies for selection of campaign parameters. Due to high computational complexity it would be difficult for larger networks.

The approach proposed in this paper assumes the generation of synthetic networks based on theoretical models, generation of network samples based on real network, performing simulations focused on verification of different seeding strategies and campaign parameters and evaluation of results with the use of MCDA methods and, finally, launching the campaign within the real network (see Fig. 1)

Simulations can be performed within synthetic networks based on theoretical models like Barabasi-Albert model (BA) [37], Watts-Strogatz (WS) model [38] and Erdos-Renyi model (ER) [39]. The size of synthetic networks can be adjusted with reference to the size of real network and it can be a fraction of the real network e.g. 10%, 20%, 30% etc. It is also important to select proper network model with high similarity to real network. The presented approach uses Kullback-Leibler measure (KL) to compare network similarities [40]. Number of nodes and edges within synthetic network can be scaled for better performance and accuracy.

Since a real network not always must be similar to idealized theoretical models, another approach can be based on network samples generated as a fraction of the real network. Snowball sampling can be used to obtain smaller structures, which would allow to perform simulations easier, yet with assumed similarity to the full network structures. Samples can be scaled from lower to higher fraction of the complete network. It is assumed that accuracy of simulations in the bigger samples is more close to the real network but computational complexity is lower for the smaller samples.

The simulations for all samples and synthetic networks are performed with the use of various campaign parameters. The number of seeds represented by the seeding fraction (SF) and its effect on total coverage can be verified and is the representation of a campaign budget. Another decisions are related to seed selection strategy (SS). It can be based on different network metrics and it is also related to campaign costs. For example, targeting high degree nodes can be more expensive than low degree nodes.

From the other point of view, the selection of nodes with high closeness can be more expensive than the selection of nodes with high degree because of higher computational

complexity required to compute closeness metrics than degree. Another tested parameters are based on propagation probabilities (PP). For lower propagation probabilities, coverage within the network will be lower, but higher probabilities require higher motivation of users to forward the content. It may require incentives and is related to increased budgets.

To compare results from samples and synthetic networks, the proposed study performs analysis for all networks used. The MCDA module takes into account possible campaign success evaluation criteria like coverage, dynamics, campaign costs. In the subsequent step, the performance table obtained from the samples, as well as the criteria and preferences, are used to produce a ranking of possible advertising strategies with the selected MCDA method. After analyzing the ranking and performing robustness / sensitivity analysis, the analyst provides the campaign parameters recommendation for real network campaign.

In the prior research [12], the authors successfully used the PROMETHEE II method [41], [42] to evaluate viral marketing campaign strategies. However, in the proposed research the authors' wanted to emphasize the effect that the marketers' weights assigned to particular criteria have on the final strategies evaluation. Therefore, it was decided that full sensitivity analysis of the obtained solutions should be performed, which eliminated aspect of uncertainty of the decision maker's criteria preference. Moreover, since in the proposed approach the input data comes from simulations, data uncertainty can also be disregarded. However, the evaluation problem at hand still is characterized by weights and data expressed on a quantitative scale. Last, but not least, the obtained solution to the strategy evaluation problem should take the form of a complete ranking to allow the choice of the best strategy. Therefore, based on the analysis of 65 MCDA methods [43], [44] and the guidelines included in [45] and [46], the authors decided to found their approach on the TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) method [47].

The TOPSIS method is a representative of the American MCDA school [48] which transforms all decision-making problem criteria into a single score value. In case of the TOPSIS method, based on the criterial performance of the evaluated criteria, a positive and negative ideal strategies are created, i.e. one which tops at each criterion and one that bottoms at all criteria. Subsequently, the score of each appraised strategy is computed as a relative distance between the strategy and both the positive and negative ideal solutions. Therefore, the best strategy would be the one which is closest to the positive ideal strategy, yet as far as possible from the negative ideal strategy in terms of criterial performance values.

#### IV. RESULTS AND DISCUSSION

##### A. Evaluation of viral marketing campaign strategies on a real network

The empirical study was based on a real network, a part of the topology of the Gnutella network as mapped in 2002 in the

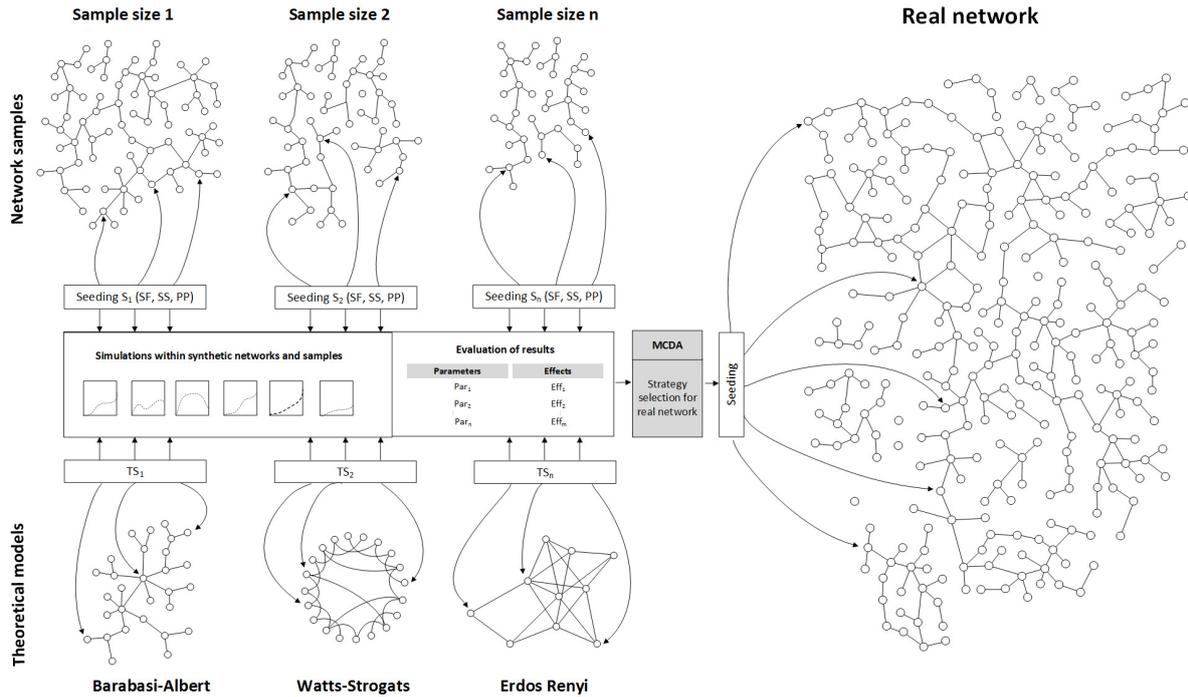


Fig. 1. Conceptual framework for real network strategy selection based on simulation results within network samples and theoretical models

[49] research. The mapped network comprises of 8846 nodes and 31839 edges. The nodes represent hosts in the Gnutella network topology and the edges represents connections between the Gnutella hosts in a single of the network snapshots collected in August 2002. The average values of the main network’s metrics are as follows:

- 1) total degree  $D = 7.1985$
- 2) closeness  $C = 1.587441e - 07$
- 3) Page Rank  $PR = 0.0001130454$
- 4) Eigen Vector  $EV = 0.01602488$
- 5) clustering coefficient  $CC = 0.0001130838$
- 6) betweenness  $B = 19104.87$

During the empirical study, the authors used the proposed framework to plan and simulate a viral marketing campaign. Ten simulation scenarios were generated to assure repeatability of the results regardless of the input parameters. Each scenario was composed of the weights drawn for each edge, ranging  $< 0; 1 >$ . These weights were later compared with the propagation probability of each node to determine whether or not the actual information propagation would occur.

As part of the simulations, a total of 400 sets of parameters were tested, built as a Cartesian product of the following simulation parameter values:

- 1) Par1 - 0.01, 0.02, ..., 0.09, 0.10;
- 2) Par2 - 0.01, 0.10, 0.20, ..., 0.90;
- 3) Par3 - degree (1), closeness (2), eigenvector centrality (3), betweenness (4) – the value is the rank of the method based on its computation speed.

Rank	Alt	CCi	SF	PP	Last Iter	Coverage Measure	
1	A11	0.7494	0.01	0.20	14.4	0.5174	1
2	A10	0.7244	0.01	0.20	14.2	0.5176	2
3	A7	0.7218	0.01	0.10	16.8	0.1334	1
4	A51	0.7148	0.02	0.20	13.2	0.5189	1
5	A50	0.7033	0.02	0.20	13.6	0.5201	2
6	A8	0.6975	0.01	0.10	19.6	0.1127	3
7	A12	0.6960	0.01	0.20	14.7	0.5172	3
8	A6	0.6901	0.01	0.10	15.6	0.1359	2
9	A47	0.6883	0.02	0.10	14.5	0.1625	1
10	A48	0.6878	0.02	0.10	18.9	0.1218	3
11	A52	0.6839	0.02	0.20	14.5	0.5181	3
12	A46	0.6805	0.02	0.10	15.1	0.1638	2
13	A91	0.6799	0.03	0.20	12.2	0.5213	1
14	A88	0.6732	0.03	0.10	18.1	0.1313	3
15	A92	0.6722	0.03	0.20	14.5	0.5194	3
16	A90	0.6635	0.03	0.20	12.4	0.5221	2
17	A128	0.6588	0.04	0.10	17.6	0.1464	3
18	A87	0.6561	0.03	0.10	13	0.1870	1
19	A131	0.6560	0.04	0.20	11.8	0.5230	1
20	A132	0.6537	0.04	0.20	14.3	0.5207	3

Fig. 2. Visualization of the top 20 alternatives from the TOPSIS evaluation of the [49] real network.

Consequently, 4000 simulations were performed for the [49] network. The results of each simulation run were registered, including inter alia the iteration during which the last infection occurred as well as the total coverage achieved, which values were labelled for the further evaluations as Eff4 and Eff5.

After the simulations concluded, the TOPSIS method was

used to evaluate all 400 campaign scenarios. Initially, the weights of all criteria were set equal. The preference direction of the **Par1-Par3** criteria was minimum and of the **Eff4-Eff5** was maximum. Intuitively that would mean the decision maker would prefer low cost of the entrepreneurship, yet long duration and maximum coverage. The top 20 strategies are presented on Fig. 2. The best strategy, A11, obtained  $\phi_{net}$  score of 0.7494. This strategy is based on low values of SF and PP (0.01 and 0.20 respectively) and degree as the method of seeding nodes selection. The runner-up alternative, A10, is based on the same SF and PP values, but uses closeness as the method for selecting the seeding nodes. As a result, slightly broader coverage was achieved in minutely less iterations (0.02s difference). The third-best strategy, A7, maintains the degree measure and the SF of 0.01, however it reduces the PP by half, to 0.10. Such strategy would non-negligibly reduce the costs of the campaign (lower investment in incentives), and, since less nodes at each step would get infected, the procedure would take longer (16.8 iterations on average). However, the obtained coverage is significantly lower, equal to 0.1334 of the network, which is over three-fold worse than the winning A11 strategy.

For the purposes of comparison, the worst strategy, A400, was based on high SF (0.10), high (ignitable) PP (0.9) and eigenvector centrality as the measure. As a result, the contamination process averagely finished within 5.1 iterations, with the mean coverage of 0.9722. Although almost full network gets covered with that strategy, it is important to note that the incentive costs for such strategy would be very high to achieve 90% propagation probability. Also the duration of the campaign would be low, which is against the DM's preferences.

One of the benefits of the TOPSIS method is the fact it allows to build an ideal reference model for the given evaluation problem. In case of the problem at hand, the ideal strategy would be based on degree for selecting the nodes to seed information to and only 1% nodes would be seeded. Incentives would be in place to generate an average propagation probability of 0.01%. With such parameters of the network, the DM would like the outcomes of the marketing campaign to be 19.6 iterations resulting in 97.22% coverage. It is important to note, however, that although ideal, such strategy is only a reference model and does not exist.

The rank presented on Fig. 2 is based on an assumption that the weight of each criterion on the final outcome is equal. However, the DM often gives more significance to some criteria over the others. One of the tremendous benefits of the utilisation of MCDA in the evaluation of viral marketing campaign strategies is the possibility to perform a sensitivity analysis, to learn how even slight changes in preferences of each criterion would affect the final outcome. Therefore, in a subsequent part of the research, a sensitivity analysis was performed to show how the ranking relations between the top 20 alternatives would change if the weights of each criterion would change. The analysis was divided into five parts, one for each criterion. During each phase, the weight of a single

criterion was changed from 1 to 100, while the weights of the remaining criteria were set equally to 50.

The results of the sensitivity analysis are presented on Fig. 3. The top row of the figure (A-E) presents how the score of each strategy changed, resulting from each criterion's weight change, whereas the bottom row of the figure (F-J) presents how that change affected the strategies' positions in the ranking. The analysis of Fig. 3A,F shows that no matter how the weight of the criterion Par1 changed, strategy A11 remained the leading one. On the other hand, if the weight of this criterion dropped slightly below 40, strategy A7 would outrun strategy A10. Strategy A51 rank is not affected by the changes of weight of criterion Par1, whilst strategy A50 (ranked fifth) would be outrun by strategy A12 (ranked 7) if its weight was higher than 75. The analysis of the chart on Fig. 3A allows to observe, that while the score of alternatives A128, A131 and A132 is not significantly affected by the changes of Par1 weight, the remaining strategies gain more score as the weight of this criterion increases. A similar tendency can be observed on Fig. 3B, where the scores of all strategies increase along with the increase of significance of criterion Par2. When the weight of that criterion would exceed 90, the runner-up strategy A7 would outrun the strategy A11. An opposite tendency can be observed on Fig. 3E, where all alternatives lose score when the weight of Eff5 grows. Along with this criterion's weight growth, there are only little changes in the order of the three leading alternatives, however, if the weight of that criterion dropped close to 0, the leading strategy A11 would drop six positions to rank 7. This demonstrates the fact that strategy A11 is considerably supported by criterion Eff5. The observation of Fig. 3F-J shows that while for the criteria Par1 and Par2 the majority of rank changes occur when the weight of the criterion changes considerably, in case of criteria Par3 – Eff5, most of the rank changes occur with even minute changes of these criteria's weights.

### B. Selection of synthetic networks

As it was presented in section IV-A, the proposed MCDA framework allows to successfully evaluate various viral marketing campaign strategies performed over a real network. However, full networks are rarely available for the entities ordering campaigns. Often, only characteristics of a network are provided. Moreover, running comprehensive simulations on a real networks containing multitude of nodes is also time consuming. Therefore, it is beneficial to perform simulations on smaller synthetic networks before launching the actual campaign on a real network.

Consequently, in the empirical research, apart from evaluating campaign strategies based on full, real network, the authors also used the proposed framework to perform simulations on synthetic networks, similar to the real one, but of a reduced size. The strategies' rankings obtained for synthetic networks were then compared to the ranking obtained for the real network.

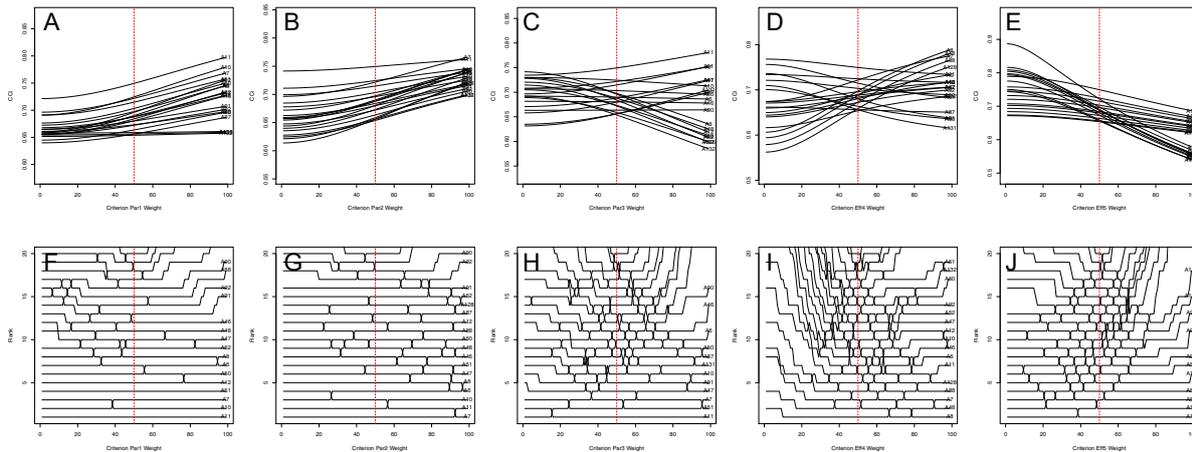


Fig. 3. Ranking sensitivity analysis for the top 20 alternatives from the TOPSIS evaluation of the [49] real network.

For the 10%, 30% and 50% size of the real network, BA, WS and ER networks were generated with the following parameters:

- 1) BA - number of nodes equal to 10%, 30% and 50% of the real network; number of edges  $m$  to add in each step equal to  $1, 2, \dots, 5$  – a total of 15 networks;
- 2) WS - number of nodes equal to 10%, 30% and 50% of the real network; the neighborhood within which the vertices of the lattice will be connected equal to  $1, 2, \dots, 5$  – a total of 15 networks;
- 3) ER - number of nodes equal to 10%, 30% and 50% of the real network; number of edges equal to the chosen number of nodes multiplied by  $1, 2, \dots, 5$  – a total of 15 networks.

As a result, a set of 45 networks was generated. In order to avoid arbitrary decisions which network to run the simulations on, the Kullback-Leibler divergence measure was used to compare the degree distribution of all generated networks to the real one. Based on the smallest value of the KLD measure, three networks were selected for further simulations. The selected networks are presented in Table IV-B.

### C. Viral marketing campaign strategies planning with synthetic networks

The results of the viral marketing campaign strategies planning with the use of the three aforementioned BA networks is presented in table on Fig. 4. The analysis of the table allows to notice that regardless of the selected network size, in all three cases the same strategy A11 was chosen as the superior one, similarly to the real network case. While in case of the real network this strategy lasted averagely for 14.4 iterations and resulted in 0.5174 coverage, in case of the synthetic networks, the process averagely lasted 10 – 11.2 iterations (slightly shorter) and resulted in 0.5783 – 0.7049 coverage (slightly higher). The second best strategy in all three synthetic networks was strategy A51, which above, in case of the real network, was ranked fourth. This strategy is based

on small values of SF and PP (0.02 and 0.20 respectively) and lasts averagely in 9.5 – 10.7 iterations resulting averagely in 0.5783 – 0.7049 coverage. The measure used here is also degree, as in the winning alternative.

The strategy A10, which for the real network evaluation was ranked second, in case of the synthetic networks reached place 3 for the 50% network and rank 4 for the remaining networks. More interesting is the case of strategy A7. On the real network it is ranked third, for the 30% network it remained at the same ranking position, however, for the 50% network it dropped to the fourth rank, whilst for the 10% network its ranking fell to 15th position. The strategy A7 is characterized by its very low SF and PP values (0.01 and 0.10 respectively) and degree as the measures which makes it one of the cheapest, with maximally extended information propagation process duration, on the cost of small final coverage. The duration of the process is very long for this strategy on the real network and the 30% and 50% networks (16.8, 12.8 and 13 iterations averagely, while the maximum average duration was 19.6, 12.9 and 13.7 iterations respectively). In case of the 10% synthetic network, the average duration is 9.7 iterations and the yielded coverage is lower, equal to 0.1784, which resulted in reduction of the A7's rank.

In case of the strategy A15 which for the 10% network is ranked third, it does not occur on the real network top-twenty list, and on the remaining synthetic networks it is below the first top-ten. This is an interesting difference, which can be further analyzed with the use of the sensitivity analysis (see Fig. 5). In case of the 10% network, the strategy is slightly supported by Par1 criterion. If the weight of criterion Par2 was increased, the strategy A15 would significantly drop in the ranking, down to rank 17. On the other hand, if the weight of the Par3 criterion became insignificant, strategy A15 would be ranked 10th. Regarding the efficiency rankings, Eff5 supports the strategy A15 (rank 11 to rank 1 increase when Eff5 weight increases from 1 to 100) and Eff4 is in conflict with A15 (rank 1 to rank 6 decrease when Eff4 weight increases from 1 to

TABLE I  
KULLBACK-LEIBLER DIVERGENCE MEASURE FOR THE SELECTED SYNTHETIC NETWORKS

Expected %	Network	Num. of nodes	Perc. of nodes	Num. of edges	Perc. of edges	KLD
10	BA, $m = 4$	885	0.100045218%	3530	0.110870316%	0.000935498
30	BA, $m = 5$	2654	0.300022609%	13255	0.416313326%	0.000800703
50	BA, $m = 5$	4423	0.5%	22100	0.694117278%	0.000521317

BA-885-4						BA-2654-5						BA-4423-5					
Rank	Alt	CCi	SF	PP	Last Iter Coverage Measure	Rank	Alt	CCi	SF	PP	Last Iter Coverage Measure	Rank	Alt	CCi	SF	PP	Last Iter Coverage Measure
1	A11	0.8202	0.01	0.20	10.3 0.5783	1	A11	0.8202	0.01	0.20	10 0.7049	1	A11	0.8190	0.01	0.20	11.2 0.7026
2	A51	0.8005	0.02	0.20	9.5 0.5783	1	A51	0.8005	0.02	0.20	9.8 0.7049	1	A51	0.7958	0.02	0.20	10.7 0.7026
3	A15	0.8002	0.01	0.30	8.7 0.8129	1	A7	0.8002	0.01	0.10	12.8 0.2567	1	A10	0.7926	0.01	0.20	11.3 0.7026
4	A10	0.7905	0.01	0.20	10.3 0.5783	2	A10	0.7905	0.01	0.20	10 0.7049	2	A7	0.7758	0.01	0.10	13 0.2686
5	A50	0.7705	0.02	0.20	9.5 0.5783	2	A91	0.7705	0.03	0.20	9.8 0.7050	1	A50	0.7679	0.02	0.20	10.7 0.7026
6	A14	0.7705	0.01	0.30	8.7 0.8129	2	A50	0.7705	0.02	0.20	9.8 0.7049	2	A91	0.7647	0.03	0.20	10.2 0.7028
7	A91	0.7689	0.03	0.20	8.8 0.5791	1	A47	0.7689	0.02	0.10	11.5 0.2732	1	A6	0.7606	0.01	0.10	13.7 0.2686
8	A55	0.7651	0.02	0.30	7.8 0.8129	1	A6	0.7651	0.01	0.10	12.8 0.2567	2	A47	0.7558	0.02	0.10	11.8 0.2831
9	A95	0.7454	0.03	0.30	7.7 0.8129	1	A131	0.7454	0.04	0.20	9.7 0.7051	1	A46	0.7430	0.02	0.10	12.4 0.2828
10	A19	0.7414	0.01	0.40	7.6 0.9173	1	A90	0.7414	0.03	0.20	9.8 0.7050	2	A131	0.7379	0.04	0.20	10 0.7033
11	A131	0.7414	0.04	0.20	8.6 0.5801	1	A46	0.7414	0.02	0.10	11.5 0.2757	2	A12	0.7372	0.01	0.20	11.3 0.7026
12	A90	0.7399	0.03	0.20	8.8 0.5786	2	A15	0.7399	0.01	0.30	8.2 0.8867	1	A90	0.7354	0.03	0.20	10.1 0.7027
13	A12	0.7369	0.01	0.20	10.6 0.5783	3	A87	0.7369	0.03	0.10	10.3 0.2898	1	A87	0.7225	0.03	0.10	10.6 0.2961
14	A54	0.7365	0.02	0.30	7.8 0.8129	2	A12	0.7365	0.01	0.20	10.1 0.7049	3	A15	0.7205	0.01	0.30	8.3 0.8882
15	A7	0.7334	0.01	0.10	9.7 0.1784	1	A8	0.7334	0.01	0.10	12.9 0.2583	3	A8	0.7193	0.01	0.10	13.5 0.2881
16	A135	0.7233	0.04	0.30	7.7 0.8129	1	A130	0.7233	0.04	0.20	9.6 0.7055	2	A52	0.7156	0.02	0.20	10.7 0.7027
17	A59	0.7205	0.02	0.40	7.2 0.9173	1	A55	0.7205	0.02	0.30	7.9 0.8867	1	A130	0.7124	0.04	0.20	10 0.7031
18	A94	0.7197	0.03	0.30	7.8 0.8129	2	A171	0.7197	0.05	0.20	9.3 0.7058	1	A55	0.7050	0.02	0.30	8.1 0.8882
19	A47	0.7184	0.02	0.10	8.8 0.2058	1	A14	0.7184	0.01	0.30	8.2 0.8867	2	A86	0.7025	0.03	0.10	10.7 0.2960
20	A52	0.7177	0.02	0.20	9.7 0.5783	3	A52	0.7177	0.02	0.20	9.8 0.7049	3	A14	0.7008	0.01	0.30	8.4 0.8882

Fig. 4. Visualization of the top 20 alternatives from the TOPSIS evaluation of the campaign strategy planning on synthetic networks.

100).

The sensitivity analysis can also provide information about the overall stability of the obtained solution. In case of the 10% network, the ranking is very stable and the A11 strategy either remains on the winning rank or drops to the second position if the weight of Par2 drops below 40%, Par3 drops below 10%, Eff4 drops below 25%. The only significant change occurs for the Eff5 criterion, where A11 would drop to rank 2 if the Eff5's weight increased to over 60% and even further if the weight increased to over 75%. If exclusively Eff5 was considered, the A11 strategy would be ranked 13th.

Similar stability for Par1-Par3 can be observed for the 30% network, however if the weight of Eff4 increased significantly or the weight of Eff5 increased significantly, A11 would be ranked 6th.

Last, but not least, in case of the 50% synthetic network, A11 would remain ranked 1st regardless of Par1 weight, would drop to 2nd position if Par2 had weight exceeding 90 or would drop to 3rd position if Par3 had negligible weight. In case of Eff4, the stability interval of the obtained solution is 0 – 80, whilst in case of Eff5 the stability interval is 35 – 100.

D. Viral marketing campaign strategies planning with network samples

As it was stated in the methodology section of this paper, although synthetic networks allow to minimize the computational efforts, their resemblance to the actual real network might be insufficient. Therefore in the subsequent step of the research, the original real network [49] was sampled, resulting in 3 networks containing 10%, 30% and 50% of the original

network. The sampling procedure was performed with the *snowball.sampling* R function from the *netdep* R library [50].

The results of the viral marketing campaign planning based on the real network [49] samples are presented in table on Fig. 6. Contrary to the synthetic networks' results, where the same strategy A11 was best in case of all three networks, in case of the samples of the real network, the rankings are more diversified.

When the 50% network is considered, the best-ranked strategy is the strategy A15, based on very low SF, higher PP (0.30), degree measure mediocre process length (14.1 iterations) and satisfying coverage (0.5075). Strategy A15 is followed by strategy A11, which uses smaller PP (0.20), which resulted in simulations in less dynamic process, leading to extending its duration to 17.9 iterations, but reducing the coverage almost by half, to 0.2685. The third position in the ranking belongs to strategy A55, which is based on 0.02 SF and 0.30 PP and results in efficiency results similar to the leading A15 strategy - 13.3 iterations and 0.5106 coverage respectively. However, the costs of such approach are higher due to the increase of the SF. When the 30% and 10% networks are considered, the A15 strategy is ranked second in the former and sixteenth in the latter, which, as mentioned earlier, is in contrast to the observations made for synthetic BA networks.

The equal-weights TOPSIS analysis was followed by a sensitivity analysis of the top 20 strategies for each of the sampled networks (see Fig. 7). An overall observation of the figures allow to see that the rankings for the 50% and 30% networks are much more stable than in case of the 10% network. To illustrate that fact, one can notice that

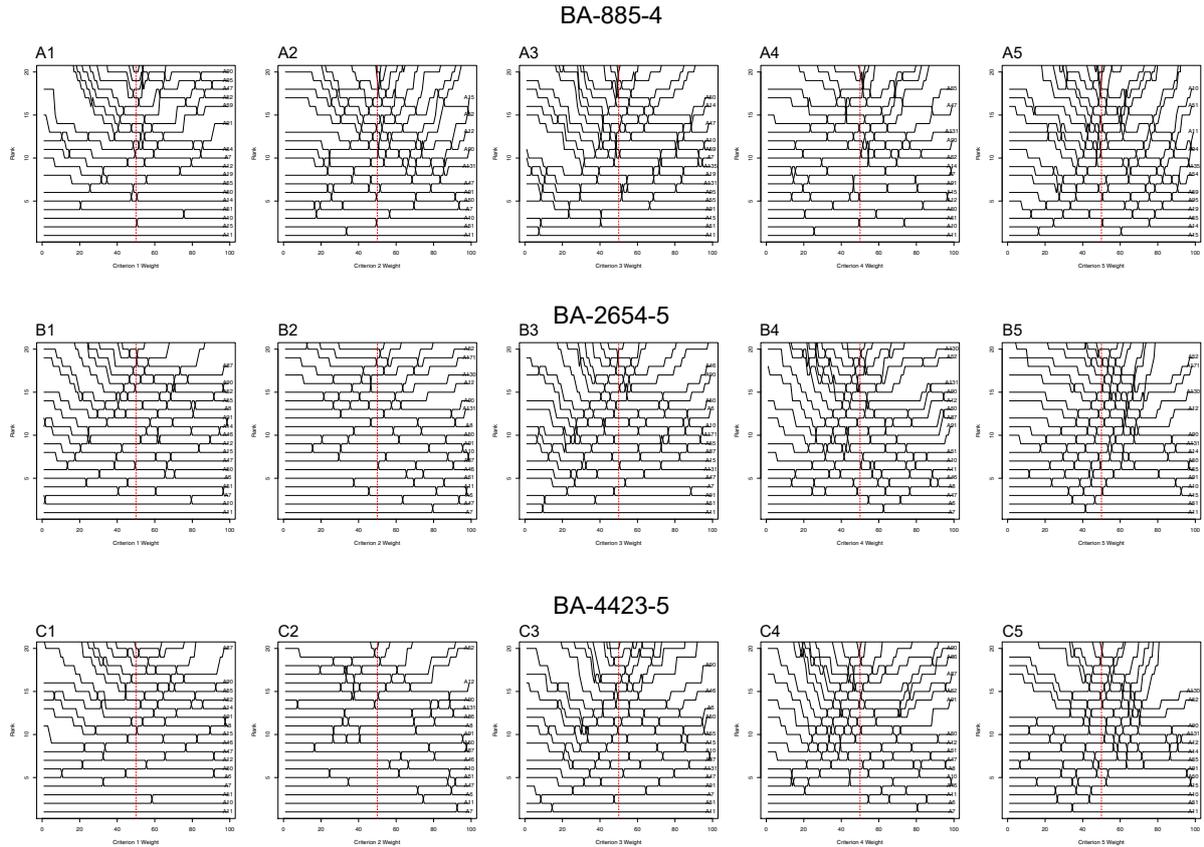


Fig. 5. Ranking sensitivity analysis for the top 20 alternatives from the TOPSIS evaluation of the synthetic networks. A1-A5 – 10% network, B1-B5 – 30% network, C1-C5 – 50% network.

in case of figures C1-C5 and B1-B5 only minute or none changes in the rank of the leading alternative can be observed when the weight of Par1-Eff5 criteria are modified. On the other hand, in case of the 10% network, if Par1 criterion weight was decreased significantly, the leading A23 strategy would drop to position 20 (see Fig. 7A1). Moreover, Fig. 7A2 and A5 demonstrate multiple leader changes in case of even slightest fluctuations of the Par2 and Eff5 criteria. When compared to the stability of the rankings obtained for the actual real network (see Fig. 3), this might suggest that a network obtained as a 10% sample of a real network is too small to maintain the stability of evaluation.

#### E. Comparison of rankings' evaluation accuracy

The research was concluded by a pairwise comparison of rankings based on equal weights for all analyzed networks. In the comparison, the scores and ranks of all strategies for each network were combined into a single table, ordered by the strategy name. This allowed to obtain correlation matrices for all the networks, presenting how correlated are the ranks (Table IV-E) and scores (Table IV-E) for each pair of networks.

The analysis of the correlation matrices allows to observe that the rankings for BA networks are highly correlated to the ranking for the real network with 0.9390 – 0.9799 correlation

coefficient for scores and 0.9631–0.9800 coefficient for ranks, which means that the relation between them is almost linear. In turn, for the sampled networks, only the ranking for the 50% network achieved high correlation coefficient with the real network, equal to 0.8797 for scores and 0.9222 for ranks. This shows, that the results of the evaluation for the real network and the 50% sampled network are very similar, yet the computational power required to perform the evaluation is significantly smaller. On the other hand, the correlation coefficient values for scores and ranks for the 30% network are much lower, i.e. 0.6043 and 0.6837 respectively, and for the 10% even lower, i.e. 0.4171 and 0.4629 respectively. Such positive yet low values of correlation coefficients indicate there is a positive relation between the rankings obtained for the real network and its 10% and 30% snowball samples. However, the margin of error there might be too high to base the actual campaign on the strategies obtained for such small network samples.

#### V. CONCLUSIONS

Nowadays, when over 45% of the world population are active social media users [51], information spreading in complex social networks begins to bring better results than traditional online advertising campaigns. Online marketers have begun

Snowball Sample 10%						Snowball Sample 30%						Snowball Sample 50%								
Rank	Alt	CCi	SF	PP	Last Iter	Coverage Measure	Alt	CCi	SF	PP	Last Iter	Coverage Measure	Alt	CCi	SF	PP	Last Iter	Coverage Measure		
1	A23	0.6298	0.01	0.50	9.1	0.1958	1	A19	0.7257	0.01	0.40	14.5	0.4418	1	A15	0.7521	0.01	0.30	14.1	0.5075
2	A27	0.6293	0.01	0.60	9.1	0.2434	1	A15	0.7137	0.01	0.30	16.3	0.2755	1	A11	0.7491	0.01	0.20	17.9	0.2685
3	A19	0.6266	0.01	0.40	9	0.1523	1	A59	0.7079	0.02	0.40	13.6	0.4448	1	A55	0.7302	0.02	0.30	13.3	0.5106
4	A22	0.6262	0.01	0.50	10.7	0.2018	2	A23	0.7003	0.01	0.50	12.7	0.5560	1	A10	0.7281	0.01	0.20	17.3	0.2748
5	A26	0.6213	0.01	0.60	10.2	0.2495	2	A18	0.6990	0.01	0.40	14.1	0.4424	2	A14	0.7258	0.01	0.30	13.9	0.5085
6	A18	0.6206	0.01	0.40	10.2	0.1590	2	A55	0.6958	0.02	0.30	14.7	0.2816	1	A51	0.7183	0.02	0.20	14.8	0.2841
7	A35	0.6182	0.01	0.80	9.4	0.3411	1	A14	0.6954	0.01	0.30	16.5	0.2781	2	A54	0.7064	0.02	0.30	13.3	0.5109
8	A31	0.6159	0.01	0.70	8.4	0.2896	1	A63	0.6832	0.02	0.50	12.1	0.5584	1	A50	0.7034	0.02	0.20	15.2	0.2883
9	A67	0.6145	0.02	0.60	8.7	0.2445	1	A99	0.6801	0.03	0.40	12.4	0.4509	1	A95	0.6982	0.03	0.30	12.3	0.5148
10	A30	0.6118	0.01	0.70	9.8	0.2959	2	A58	0.6792	0.02	0.40	13.2	0.4461	2	A91	0.6965	0.03	0.20	13.8	0.3017
11	A39	0.6101	0.01	0.90	10.1	0.3828	1	A95	0.6765	0.03	0.30	13.6	0.2963	1	A12	0.6912	0.01	0.20	17.3	0.2702
12	A34	0.6084	0.01	0.80	10.2	0.3485	2	A22	0.6739	0.01	0.50	12.4	0.5570	2	A16	0.6908	0.01	0.30	14.4	0.5077
13	A62	0.6084	0.02	0.50	9.6	0.2087	2	A27	0.6661	0.01	0.60	11.5	0.6388	1	A90	0.6835	0.03	0.20	14.4	0.3022
14	A14	0.6083	0.01	0.30	10.2	0.1050	2	A20	0.6658	0.01	0.40	14.8	0.4409	3	A52	0.6813	0.02	0.20	16.9	0.2762
15	A63	0.6079	0.02	0.50	8.1	0.1974	1	A103	0.6631	0.03	0.50	11.6	0.5624	1	A56	0.6744	0.02	0.30	13.9	0.5104
16	A15	0.6062	0.01	0.30	8	0.1031	1	A54	0.6609	0.02	0.30	13.4	0.2851	2	A19	0.6723	0.01	0.40	10.2	0.6304
17	A66	0.6044	0.02	0.60	9.4	0.2547	2	A139	0.6597	0.04	0.40	12.1	0.4573	1	A131	0.6723	0.04	0.20	13.1	0.3126
18	A75	0.6041	0.02	0.80	8.9	0.3423	1	A16	0.6588	0.01	0.30	16.4	0.2732	3	A94	0.6721	0.03	0.30	12.2	0.5151
19	A107	0.6015	0.03	0.60	8.4	0.2527	1	A62	0.6582	0.02	0.50	12	0.5598	2	A135	0.6715	0.04	0.30	11.8	0.5174
20	A102	0.5998	0.03	0.50	9.5	0.2207	2	A98	0.6507	0.03	0.40	12.1	0.4523	2	A92	0.6665	0.03	0.20	16.4	0.2825

Fig. 6. Visualization of the top 20 alternatives from the TOPSIS evaluation of the campaign strategy planning on the real network [49] samples.

TABLE II  
CORRELATION MATRIX BETWEEN THE RANKS OF EACH OF THE ANALYZED NETWORKS.

Rank	Real	BA-885-4	BA-2654-5	BA-4423-5	SS 10%	SS 30%	SS 50%
Real	x	0.9631	0.9794	0.9800	0.4629	0.6837	0.9222
BA-885-4	0.9631	x	0.9840	0.9812	0.4806	0.7760	0.9703
BA-2654-5	0.9794	0.9840	x	0.9980	0.3809	0.6706	0.9289
BA-4423-5	0.9800	0.9812	0.9980	x	0.3647	0.6585	0.9191
SS 10%	0.4629	0.4806	0.3809	0.3647	x	0.8227	0.6159
SS 30%	0.6837	0.7760	0.6706	0.6585	0.8227	x	0.8718
SS 50%	0.9222	0.9703	0.9289	0.9191	0.6159	0.8718	x

TABLE III  
CORRELATION MATRIX BETWEEN THE SCORE VALUES OF EACH OF THE ANALYZED NETWORKS.

CCi	Real	BA-885-4	BA-2654-5	BA-4423-5	SS 10%	SS 30%	SS 50%
Real	x	0.9390	0.9749	0.9799	0.4171	0.6043	0.8797
BA-885-4	0.9390	x	0.9757	0.9729	0.4954	0.7730	0.9688
BA-2654-5	0.9749	0.9757	x	0.9974	0.3807	0.6373	0.9152
BA-4423-5	0.9799	0.9729	0.9974	x	0.3674	0.6266	0.9049
SS 10%	0.4171	0.4954	0.3807	0.3674	x	0.8204	0.6043
SS 30%	0.6043	0.7730	0.6373	0.6266	0.8204	x	0.8480
SS 50%	0.8797	0.9688	0.9152	0.9049	0.6043	0.8480	x

to invest greater effort into seeding information into social networks and providing incentives to increase the information propagation probability within the networks. These increased efforts have opened the research area for providing evaluation of various social network advertising campaign strategies as well as supporting the process of their planning.

The approach presented in this paper provides a framework for multi-criteria planning of viral marketing campaigns in social networks and their evaluation, in which various preferences and criteria of the marketer are taken into account. The example criteria provided in this paper allow to choose the satisfactory campaign strategy considering the costs related to the seeding of the information and providing incentives to

increase its propagation probability in relation to their effect on the process dynamics and obtained coverage.

The authors' contributions in this paper include:

- multi-criteria framework for evaluation of viral marketing campaigns in social networks;
- simulation engine and usage of synthetic network models and real network samples of limited size allowed to provide a viral marketing campaigns planning tool of reduced computational requirements;
- an example set of criteria was provided that allows to choose a satisfactory viral marketing campaign strategy based on multi-criteria consideration of its costs, dynamics and coverage;

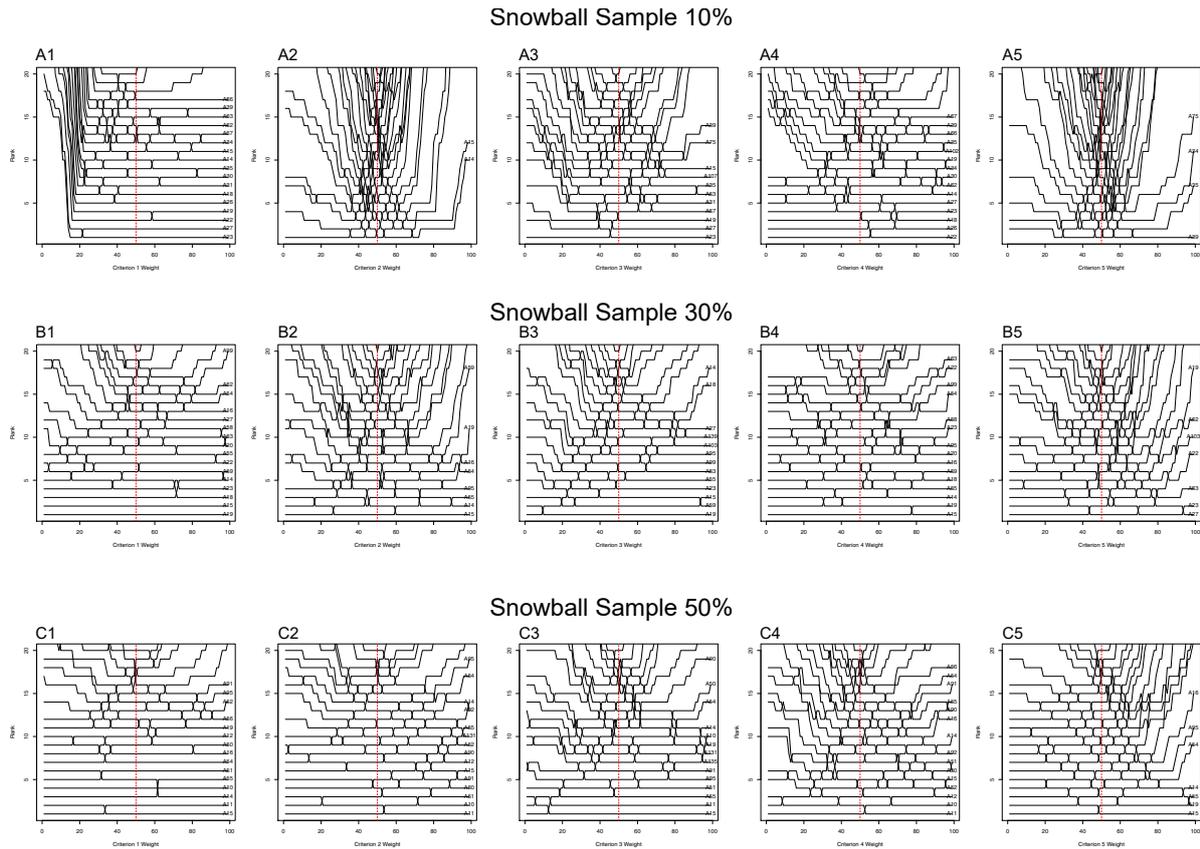


Fig. 7. Ranking sensitivity analysis for the top 20 alternatives from the TOPSIS evaluation of the real network [49] samples. A1-A5 – 10% network, B1-B5 – 30% network, C1-C5 – 50% network.

- the strategies' evaluation accuracy was compared between a full-size real network and a set of reduced-size synthetic and sample networks derived from the original network.

In practical terms, the empirical study has shown that while the synthetic networks, which were selected based on their Kullback-Leibler divergence, provided very similar results to the real network even when as little as 10% of nodes were used, in case of the sampled networks obtained with the snowball sampling approach provided satisfactory results only when the number of nodes was still relatively high. Also, while the rankings obtained from synthetic networks were stable, there was little stability of the rankings from the snowball sample networks.

All in all, the research has identified possible areas of improvement and future works. First of all, a more numerous set of sizes of sample network could be studied to verify how the network size affects its rankings' correlation to the real network's rankings. Secondly, only snowball sampling approach was used in the research. It would be beneficial to explore networks obtained with other sampling approaches. Last, but not least, the list of criteria could be expanded to allow more precise adjustment of the selected strategy to the marketer's needs.

## VI. ACKNOWLEDGMENTS

This work was supported by the National Science Centre, Poland, grant no. 2016/21/B/HS4/01562.

## REFERENCES

- [1] W. Chmielarz and O. Szumski, "Digital distribution of video games—an empirical study of game distribution platforms from the perspective of polish students (future managers)," in *Information Technology for Management: Emerging Research and Applications*. Springer, 2018, pp. 136–154.
- [2] D. J. Watts, J. Peretti, and M. Frumin, *Viral marketing for the real world*. Harvard Business School Pub., 2007.
- [3] E. Ziemba, *Towards a sustainable information society: People, business and public administration perspectives*. Cambridge Scholars Publishing, 2016.
- [4] K. Szopik-Depczynska, A. Kedzierska-Szczepaniak, K. Szczepaniak, K. Cheba, W. Gajda, and G. Ioppolo, "Innovation in sustainable development: an investigation of the EU context using 2030 agenda indicators," *Land Use Policy*, vol. 79, pp. 251–262, Dec. 2018. doi: 10.1016/j.landusepol.2018.08.004. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0264837718306203>
- [5] W. Chmielarz and O. Szumski, "Analysis of users of computer games," in *2016 Federated Conference on Computer Science and Information Systems (FedCSIS)*. IEEE, 2016, pp. 1139–1146.
- [6] O. Hinz, B. Skiera, C. Barrot, and J. U. Becker, "Seeding strategies for viral marketing: An empirical comparison," *Journal of Marketing*, vol. 75, no. 6, pp. 55–71, 2011.

- [7] J. Tang, M. Musolesi, C. Mascolo, V. Latora, and V. Nicosia, "Analysing information flows and key mediators through temporal centrality metrics," in *Proceedings of the 3rd Workshop on Social Network Systems*. ACM, 2010, p. 3.
- [8] M. Salehi, R. Sharma, M. Marzolla, M. Magnani, P. Siyari, and D. Montesi, "Spreading processes in multilayer networks," *IEEE Transactions on Network Science and Engineering*, vol. 2, no. 2, pp. 65–83, 2015.
- [9] K. Kandhway and J. Kuri, "How to run a campaign: Optimal control of sis and sir information epidemics," *Applied Mathematics and Computation*, vol. 231, pp. 79–92, 2014.
- [10] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2003, pp. 137–146.
- [11] E. M. Rogers, *Diffusion of innovations*. Simon and Schuster, 2010.
- [12] A. Karczmarczyk, J. Jankowski, and J. Wątróbski, "Multi-criteria decision support for planning and evaluation of performance of viral marketing campaigns in social networks," *PLOS ONE*, vol. 13, no. 12, p. e0209372, Dec. 2018. doi: 10.1371/journal.pone.0209372. [Online]. Available: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0209372>
- [13] E. Ziemba, "The contribution of ict adoption to the sustainable information society," *Journal of Computer Information Systems*, vol. 59, no. 2, pp. 116–126, 2019.
- [14] G. Bello-Orgaz, J. J. Jung, and D. Camacho, "Social big data: Recent achievements and new challenges," *Information Fusion*, vol. 28, pp. 45–59, 2016.
- [15] J. Wątróbski, E. Ziemba, A. Karczmarczyk, and J. Jankowski, "An index to measure the sustainable information society: the polish households case," *Sustainability*, vol. 10, no. 9, p. 3223, 2018.
- [16] J. Jankowski, J. Hamari, and J. Wątróbski, "A gradual approach for maximising user conversion without compromising experience with high visual intensity website elements," *Internet Research*, vol. 29, no. 1, pp. 194–217, 2019.
- [17] R. Pfitzner, A. Garas, and F. Schweitzer, "Emotional divergence influences information spreading in twitter," *ICWSM*, vol. 12, pp. 2–5, 2012.
- [18] R. Michalski, T. Kajdanowicz, P. Bródka, and P. Kazienko, "Seed selection for spread of influence in social networks: Temporal vs. static approach," *New Generation Computing*, vol. 32, no. 3-4, pp. 213–235, 2014.
- [19] C. Kiss and M. Bichler, "Identification of influencers: measuring influence in customer networks," *Decision Support Systems*, vol. 46, no. 1, pp. 233–253, 2008.
- [20] Y. Liu-Thompkins, "Seeding viral content: The role of message and network factors," *Journal of Advertising Research*, vol. 52, no. 4, pp. 465–478, 2012.
- [21] L. Seeman and Y. Singer, "Adaptive seeding in social networks," in *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on*. IEEE, 2013, pp. 459–468.
- [22] J. Jankowski, P. Bródka, P. Kazienko, B. K. Szymanski, R. Michalski, and T. Kajdanowicz, "Balancing speed and coverage by sequential seeding in complex networks," *Scientific reports*, vol. 7, no. 1, p. 891, 2017.
- [23] J. Jankowski, "Dynamic rankings for seed selection in complex networks: Balancing costs and coverage," *Entropy*, vol. 19, no. 4, p. 170, 2017.
- [24] J.-L. He, Y. Fu, and D.-B. Chen, "A novel top-k strategy for influence maximization in complex networks with community structure," *PloS one*, vol. 10, no. 12, p. e0145283, 2015.
- [25] J.-X. Zhang, D.-B. Chen, Q. Dong, and Z.-D. Zhao, "Identifying a set of influential spreaders in complex networks," *Scientific reports*, vol. 6, p. 27823, 2016.
- [26] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, "Identification of influential spreaders in complex networks," *Nature physics*, vol. 6, no. 11, p. 888, 2010.
- [27] C. Granell, S. Gómez, and A. Arenas, "Competing spreading processes on multiplex networks: awareness and epidemics," *Physical review E*, vol. 90, no. 1, p. 012808, 2014.
- [28] C. Granell, S. Gómez, and A. Arenas, "Dynamical interplay between awareness and epidemic spreading in multiplex networks," *Physical review letters*, vol. 111, no. 12, p. 128701, 2013.
- [29] X. Wei, N. C. Valler, B. A. Prakash, I. Neamtii, M. Faloutsos, and C. Faloutsos, "Competing memes propagation on networks: A network science perspective," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 6, pp. 1049–1060, 2013.
- [30] M. Bampo, M. T. Ewing, D. R. Mather, D. Stewart, and M. Wallace, "The effects of the social structure of digital networks on viral marketing performance," *Information systems research*, vol. 19, no. 3, pp. 273–290, 2008.
- [31] J. Y. Ho and M. Dempsey, "Viral marketing: Motivations to forward online content," *Journal of Business research*, vol. 63, no. 9-10, pp. 1000–1006, 2010.
- [32] S. Stieglitz and L. Dang-Xuan, "Emotions and information diffusion in social media: sentiment of microblogs and sharing behavior," *Journal of management information systems*, vol. 29, no. 4, pp. 217–248, 2013.
- [33] A. Dobeles, A. Lindgreen, M. Beverland, J. Vanhamme, and R. Van Wijk, "Why pass on viral messages? because they connect emotionally," *Business Horizons*, vol. 50, no. 4, pp. 291–304, 2007.
- [34] C. Camarero and R. San José, "Social and attitudinal determinants of viral marketing dynamics," *Computers in Human Behavior*, vol. 27, no. 6, pp. 2292–2300, 2011.
- [35] J. Berger and K. L. Milkman, "What makes online content viral?" *Journal of marketing research*, vol. 49, no. 2, pp. 192–205, 2012.
- [36] A. Rezvani, B. Moradabadi, M. Ghavipour, M. M. D. Khomami, and M. R. Meybodi, "Social network sampling," in *Learning Automata Approach for Social Networks*. Springer, 2019, pp. 91–149.
- [37] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Reviews of modern physics*, vol. 74, no. 1, p. 47, 2002.
- [38] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *nature*, vol. 393, no. 6684, p. 440, 1998.
- [39] P. Erdős and A. Rényi, "On random graphs, i," *Publicationes Mathematicae (Debrecen)*, vol. 6, pp. 290–297, 1959.
- [40] S. Kullback and R. A. Leibler, "On information and sufficiency," *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951.
- [41] J. Wątróbski, K. Małeck, K. Kijewska, S. Iwan, A. Karczmarczyk, and R. Thompson, "Multi-criteria analysis of electric vans for city logistics," *Sustainability*, vol. 9, no. 8, p. 1453, 2017.
- [42] P. Ziemba, "Neat f-promethee—a new fuzzy multiple criteria decision making method based on the adjustment of mapping trapezoidal fuzzy numbers," *Expert Systems with Applications*, vol. 110, pp. 363–380, 2018.
- [43] J. Wątróbski, J. Jankowski, P. Ziemba, A. Karczmarczyk, and M. Ziolo, "Generalised framework for multi-criteria method selection," *Omega*, vol. 86, pp. 107–124, 2019.
- [44] J. Wątróbski, J. Jankowski, P. Ziemba, A. Karczmarczyk, and M. Ziolo, "Generalised framework for multi-criteria method selection: Rule set database and exemplary decision support system implementation blueprints," *Data in brief*, vol. 22, p. 639, 2019.
- [45] J. Wątróbski, J. Jankowski, and Z. Piotrowski, "The selection of multicriteria method based on unstructured decision problem description," in *International Conference on Computational Collective Intelligence*. Springer, 2014, pp. 454–465.
- [46] J. Wątróbski and J. Jankowski, "Guideline for media method selection in production management area," in *New frontiers in information and production systems modelling and analysis*. Springer, 2016, pp. 119–138.
- [47] W. Chmielarz and M. Zborowski, "Analysis of e-banking websites' quality with the application of the topsis method—a practical study," *Procedia computer science*, vol. 126, pp. 1964–1976, 2018.
- [48] M. Jankowski, A. Borsukiewicz, K. Szopik-Depczynska, and G. Ioppolo, "Determination of an optimal pinch point temperature difference interval in ORC power plant using multi-objective approach," *Journal of Cleaner Production*, vol. 217, pp. 798–807, Apr. 2019. doi: 10.1016/j.jclepro.2019.01.250. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0959652619302756>
- [49] M. Ripeanu, I. Foster, and A. Iamnitchi, "Mapping the Gnutella Network: Properties of Large-Scale Peer-to-Peer Systems and Implications for System Design," *arXiv:cs/0209028*, Sep. 2002, arXiv: cs/0209028. [Online]. Available: <http://arxiv.org/abs/cs/0209028>
- [50] "Snowball Sampling Function - R Documentation." [Online]. Available: <https://www.rdocumentation.org/packages/netdep/versions/0.1.0/topics/snowball.sampling>
- [51] S. Kemp, "Digital 2019: Global Internet Use Accelerates," Jan. 2019. [Online]. Available: <https://wearesocial.com/blog/2019/01/digital-2019-global-internet-use-accelerates>



## An Approach to Customer Community Discovery

Jerzy Korczak

International University of  
Logistics and Transport  
ul. Sołtysowicka 19B  
51-168 Wrocław, Poland  
Email: jerzy.korczak@ue.wroc.pl

Maciej Pondel

Wrocław University of Economics,  
ul. Komandorska 118-120  
53-345 Wrocław, Poland  
Email: maciej.pondel@ue.wroc.pl

Wiktor Sroka

Wrocław University of Economics,  
ul. Komandorska 118-120  
53-345 Wrocław, Poland  
Assentis Technologies AG,  
Blegistrasse 1, 6343 Rotkreuz,  
Switzerland  
Email: wiktorsroka@gmail.com

□ *Abstract*—In the paper, a new multi-level hybrid method of community detection combining a density-based clustering with a label propagation method is proposed. Many algorithms have been applied to preprocess, visualize, cluster, and interpret the data describing customer behavior, among others DBSCAN, RFM, k-NN, UMAP, LPA. In the paper, two key algorithms have been detailed: DBSCAN and LPA. DBSCAN is a density-based clustering algorithm. However, managers usually find the clustering results too difficult to interpret and apply. To enhance the business value of clustering and create customer communities, the label propagation algorithm (LPA) has been proposed due to its quality and low computational complexity. The approach is validated on real life marketing database using advanced analytics platform Upsaily.

### I. INTRODUCTION

DETECTING communities is one of the usual and important problems in modern data analysis of decision support systems. Many approaches and algorithms of community discovery have been published in network literature [1]-[7]. A community can be considered as a densely connected group of nodes that is only loosely connected to the rest of the network [8]. An example of such a community in a large network is a set of customers in marketing information systems having similar profile or behavior [9].

In recent years, the efficient data mining of large volume and high dimensional data has become of utmost importance. Therefore, applying the most appropriate method of obtaining accurate and business-oriented partitions is crucial. In literature one can find many clustering algorithms, starting with classical k-means, through density-based, partitioning, self-organizing maps, graph-based, grid-based, to combinational and hybrid solutions. These algorithms are usually evaluated based on clustering measurements, showing that some clustering techniques are better for large datasets while others give good results finding clusters with arbitrary shapes. Nonetheless, there is no one algorithm that can achieve the

best performance on all measurements for any given dataset [4][10][11][13] and also obtain the best results.

Therefore, in marketing analysis, discovering accurate and business focused partitions using a single algorithm in isolation becomes highly complex. There are many reasons for these difficulties: sensitivity to initial values, unknown quantity of expected clusters, non-spherical datasets, sensitivity to noise and outliers, varying densities of clusters, or difficulties of business interpretation.

To strengthen the business outcomes and reduce weaknesses of the single algorithm approaches, a new hybrid multi-level method of community discovery will be proposed. It combines density-based clustering with business-oriented label propagation method. Five basic algorithms have been integrated into this method: DBSCAN, RFM, k-NN, UMAP and LPA. The DBSCAN, which has already been used in many applications [10]-[13], is taken as the density-based algorithms. DBSCAN identifies clusters by measuring density as the number of observations in a designated area. If the density is greater than the density of observations belonging to other clusters, then the defined area is identified as a cluster. Usually, in business application, DBSCAN creates a lot of difficult to interpret clusters. To improve cluster quality and interpretation, a second algorithm is proposed that enriches the results of DBSCAN and is able to form communities. After analysis of various community detection methods, the label propagation algorithm (LPA) was selected due to its simplicity and low computational complexity. The LPA was proposed by Raghavan et al. [14] for detecting communities in large networks. The idea of label propagation is as follows: before beginning computation, some nodes of the network possess assigned labels. During process execution, the labels are propagated iteratively throughout the network according to the formula below.

$$g_j = \arg \max_g \sum_j A_{ij} \delta(g_j g) \quad (1)$$

where  $A_{ij}$  is an element of the adjacency matrix of the network,  $\delta$  is equal to 1 when its arguments are the same, and 0 otherwise. There are many extensions of original label propagation algorithm [15], [8], [16]. In our approach, a weighted network is assumed, so formula (1) is rewritten as:

□ The project was partially financed by the Ministry of Science and Higher Education in Poland under the programme "Regional Initiative of Excellence" 2019 - 2022 project number 015/RID/2018/19.

$$g_j = \arg \max_g \sum_j W_{ij} \delta(g_j, g) \quad (2)$$

where  $W_{ij}$  is the sum of weights on the edges between nodes  $i$  and  $j$  of the adjacency matrix of the network,  $\delta$  is equal to 1 when its arguments are the same, and 0 otherwise.

In other words, the nodes sequentially adopt the labels shared by most of their neighbors taking into consideration the weights of the edges. The propagation ends when the labels no longer change.

It is important to note that in our case study nodes are represented by clusters of customers created by DBSCAN. Neighborhoods of clusters are defined individually by the distance between the centers of clusters. The upper limit of neighboring is usually predefined by the manager or analysts, so the number of neighboring clusters is variable.

The business goal of the study is to obtain a higher quality of definition of customer communities from the marketing viewpoint. Therefore, in the approach the Recency Frequency Monetary value method has been integrated with graph clustering to give clusters of higher quality compared to the traditional mono-algorithm clustering. Various data sources, different quality measures, and business orientation provide more up-to-date and richer information for decision makers and marketing analysts.

The paper is structured as follows: in the next section, the basic characteristics of customers profiles and behavior are provided. The information is saved in the database and available using Upsaily platform. The third section describes a method of clustering of customers of the internet store and the measures to evaluate quality of the results. The fourth section details the label propagation algorithm and a method of discovery of customer communities. The results of the case study on real life database are presented and discussed in the last section. A general conclusion summarizes the outcomes of the proposed approach.

## II. ANALYSIS OF CUSTOMER BEHAVIOR AND PROPERTIES

The first studies of the subject of customer behavior were conducted more than 60 years ago [17], [18]. They focused on customer identification in offline stores, analysis of customer characteristics, and studies on buying-behavior patterns. It is quite common to find customer-behavior research based on questionnaires filled by researcher and a customer who would have accepted to take part in such a study [19]-[21]. Such research is time- and resource-consuming; however, but what is more important is the fact that customers behave differently when they are aware of participation in research.

Currently, analysis of customer behavior in e-commerce is much more convenient and more options can be applied. It is possible as today's e-commerce databases collect data about every single action the customer undertakes (visit, transaction, search, and many more) [17], [19]. Such systems concentrate on a delivery of the best fitting proposal for a customer in a perspective of the selected customer

segment, desired product, and conditions under which the product is offered. Those issues were examined by the authors in [12] using customer clustering based on the RFM method, considering customer recency, frequency of purchases, and monetary value of orders. RFM method has been shown to be very useful in determining the proper point in time to provide customer with an offer.

This paper concentrates on another set of characteristics describing customer behavior. The proposed segmentation was inspired by direct interviews with e-commerce managers who independently observed two principals in terms of profit generation, but also contrary segments of customers. One of the segments brings together fashion-driven customers (they focus on new and fashionable items). Second one is "bargain hunters" – discount-driven customers who are ready to purchase products present on a market for a longer period of time. This segmentation is called "fashion vs. discount".

Being aware of such an observation, the authors extracted data from transactional database of the structure presented in Fig. 1. Due to a large number of tables and attributes in the source database, only tables and fields used in the experiment are presented. The data used in the experiment come from a fashion store (clothes and related items).

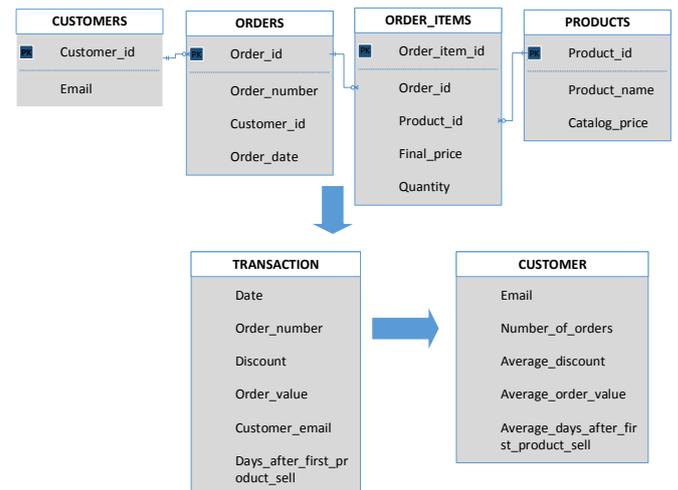


Fig. 1. Structure of source database and result tables

Having such a source database, the following measures characterizing transactions are computed (TRANSACTION table in Fig. 1):

- Value – as a quantity of items multiplied by price.
- Discount – as a percentage the difference between the highest transactional price of a specific item and its price in the current transaction.
- Days after first item sell – as the number of days from the first transaction of a given item and the current transaction.

In order to build customer profile (CUSTOMER table in Fig. 1): the above measures have been aggregated to define:

- Loyalty – expressed in the number of orders. Such an attribute differentiates the one-off buyer customer from the loyal customers.

- Average discount – high percentage discounts are typical for bargain hunters.
- Average number of days after first product sell – determines whether the customer is interested in new (fashionable) items or accepts purchasing items launched in previous seasons.
- Average order value – determines the amount of money the customer is able to spend for a single purchase.

Sample data being the input to the experiment is presented in Fig. 2. Whole data set included 264127 rows (customers).

email	average days after first product sell	average discount	average order value	number of orders
00000b3a11d76 added26b0	204,88	44	183,00	8,00
490c4eb0bdaf@unity.pl				
00005bad570278ac7739	203,25	8	244,00	1,00
9df05c96ebf1@unity.pl				
0000739e973086436944	224,24	24	294,00	9,00
624a10acb44a@unity.pl				
000084e2f4cceb412cc	245,63	57	158,00	2,00
c3bb8270e20@unity.pl				
0000aaab4a6e7bec76ba	114,00	33	103,00	1,00
668d507c0b9d@unity.pl				

Fig. 2 Sample input data

The Upsaily platform, directed to retail companies working in both B2C and B2B models, is geared towards current customers of the online internet shops. The experiment presented in this paper is based on database originating from B2C store. In the database, not only all customer transactions are stored (which is presented in Fig. 1), but also the basic data about their demographic and behavioral profile. Functionally, the solution can be classified as a Customer Intelligence system, i.e. one whose primary interest is current customers. The aim is not to help in acquiring new customers, but to increase customer satisfaction that translates into increasing turnover. It can be achieved by customers making follow-up purchases, increasing the value of individual orders (cross-selling) or more valuable products (up-selling). The Customer Intelligence approach is related to conducting analytical activities leading to creation of a clear image of the customer so that one can find the most valuable customers and send them a personalized marketing message. The system is equipped with several analyses including customer segmentation. The main screen of the platform where a manager can search for desired analysis is presented in Fig. 3.

The multi-level approach to discover customer communities will be described in the following steps:

1. Customers clustering using HDBSCAN algorithm.
2. Dimensions reduction using Uniform Manifold Approximation and Projection (UMAP) method in order to base next steps on two dimensions.
3. Centroid calculation for each cluster according to UMAP result.
4. Graph generation with k-NN algorithm.
5. Communities detection according to LPA algorithm.
6. Marketing interpretation.

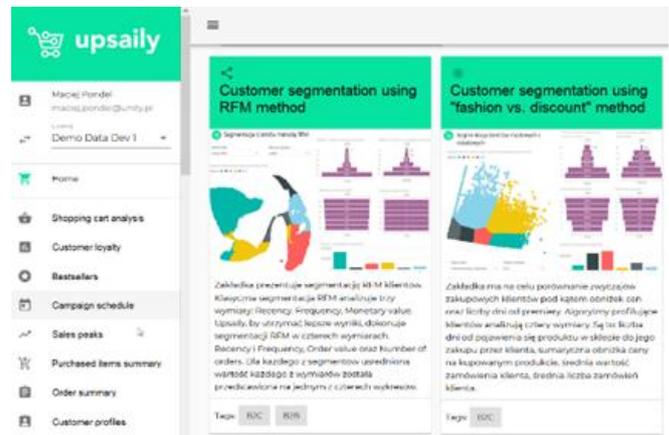


Fig. 3 Main screen of Upsaily platform

Details on the particular steps will be given in the next sections.

### III. CLUSTERING OF CUSTOMERS OF THE INTERNET STORE

Upsaily platform is equipped with two main business customer segmentations based on RFM and the “fashion vs. discount” method explained earlier. Upsaily uses four algorithms, namely:

- k-means based on the Euclidean distance between observations.
- Bisecting k-means acting on a basis similar to k-means, however, starting with all the observations in one cluster and then dividing the cluster into two sub-clusters, using the k-means algorithm.
- Gaussian Mixture Model (GMM), which is a probabilistic model based on the assumption that a particular feature has a finite number of normal distributions.
- HDBSCAN which is an extension of DBSCAN algorithm presented in [22]. The original DBSCAN identifies clusters by measuring density as the number of observations in a designated area. If the density is greater than the density of observations belonging to other clusters, then the defined area is identified as a cluster.

Experiments with each algorithm indicating their strengths and weaknesses as well as collaborative approaches have already been presented in [12].

In the current experiment, we would like to identify small but very precise segments of the most profitable customers. A profitable customer is one whose order values are high and at the same time they don't seek discounts. Authors have done corresponding clustering using k-means clustering algorithm in order to evaluate proposed method by comparison with the typical approach. Source data included 264127 rows describing customers (presented in Fig. 2). 140 segments were generated. Fig. 4 presents visualization of 7 selected clusters of the most profitable customers. Customers assigned to clusters (indicated by color) are presented in left hand side. Distribution of order values is presented in right hand side.

Methods of clustering assessment were presented in [12], [23]. For the interesting clusters, the measure of scatter within the cluster using the Davies-Bouldin index is computed. In general, the lower the value of the measure, the more consistent a cluster is. In this experiment measure of scatter was between 38.68 (best value) and 168.01 (worst value). Average value in seven selected clusters is 80.47.

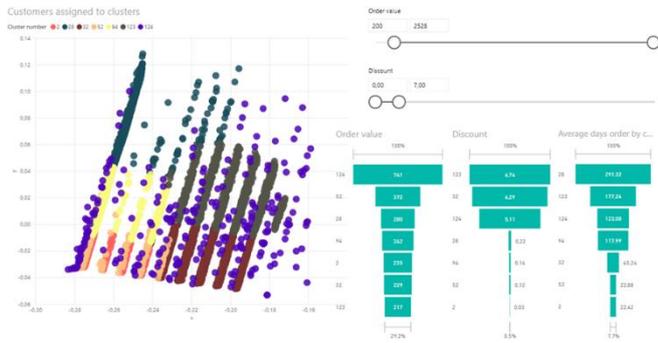


Fig. 4 Most profitable customers in k-means segmentation

In order to perform current experiment, HDBSCAN method was selected because of its marketing usage in effective discovery of clear patterns in given set of observations. It is worth mentioning that other algorithms are focused on assigning observations to a specific number of clusters defined by user upfront. HDBSCAN generates the number of clusters based on the number of patterns found in the data. It can also leave some observations without assigning them to any cluster if no pattern is found.

To understand the idea of HDBSCAN, the basic DBSCAN has to be explained.

The algorithm can be abstracted also into the following steps [24]:

1. Find the points in the  $\epsilon$ -neighbourhood of every point and identify the core points with more than  $\text{minPts}$  neighbours.
2. Find the connected components (subgraph) of core points on the neighbor graph, ignoring all non-core points.
3. Assign each non-core point to a nearby cluster if the cluster is an  $\epsilon$ -neighbor, otherwise assign it to noise (outliers).

The DBSCAN algorithm can be parameterized by  $\epsilon$  (eps) defining the minimum distance between two points and  $\text{minPts}$  denoting the minimum number of points to form a dense region.

DBSCAN algorithm in pseudo code is given [25]

```
DBSCAN(SetOfPoints, Eps, MinPts)
//SetOfPoints is UNCLASSIFIED
ClusterId := nextId(NOISE);
FOR i FROM 1 TO SetOfPoints.size DO
  Point := SetOfPoints.get(i);
  IF Point.CiId = UNCLASSIFIED THEN
    IF
      ExpandCluster(SetOfPoints, Point, ClusterId, Eps, MinPts) THEN
        ClusterId := nextId(ClusterId);
```

```
END IF;
END IF;
END FOR;
END; // DBSCAN
```

ExpandCluster function checks all points in neighbourhood of a given point. If number of those points is higher than  $\text{minPts}$  parameter, they are assigned to cluster otherwise to noise.

HDBSCAN converts DBSCAN into a hierarchical clustering algorithm, and then uses a technique to extract a flat clustering based on the stability of clusters. The following steps present the idea of HDBSCAN [22]:

1. Transform the space according to the density/ sparsity.
2. Build the minimum spanning tree of the distance weighted graph.
3. Construct a cluster hierarchy of connected components.
4. Condense the cluster hierarchy based on minimum cluster size.
5. Extract the stable clusters from the condensed tree.

The result of step 1, customers assigned to clusters using HDBSCAN algorithm, is presented in Fig. 5.

email	Cluster number
00005bad570278ac77399df05c96ebf1@unity.pl	562
0000739e973086436944624a10acb44a@unity.pl	-1
000084e2f4cceb412ccc3bb8270e20@unity.pl	1428
0000aaab4a6e7bec76ba668d507c0b9d@unity.pl	2001
0000c55228e230f7328cb8d30cfa6a56@unity.pl	1126
0000ec3ee95d687941ffdb40a2978e10@unity.pl	1454
000106e34137ad831fcadb9f1a136c30@unity.pl	384
00014afc8bd6d898b44829fe59bf4a3a@unity.pl	1570
000230f75c317ad9c09cf9fcaa885f9d@unity.pl	445
00027aefbaecbdf9e12a325befd2e9bf@unity.pl	1999
0002c0d61e903bbe749507093563f03b@unity.pl	1207
0003067477d1c77200568dadcf8728cc@unity.pl	1447
00039e775949311810649e8ec21359b2@unity.pl	445
0003b08584f2cb8e3bce38a7750a6590@unity.pl	-1
0003d34a40ef266469c39697157f89ba@unity.pl	707

Fig. 5 Customers assigned to HDBSCAN generated clusters

The result of step 2 and 3 which concerns a dimension reduction using UMAP method [26] and centroid calculation is presented in Fig. 6. Centroids are described by  $x$  and  $y$  coordinates.

Cluster number	x	y	Number of customers
-1	7,39	0,13	89830
0	13,06	4,62	126
1	21,54	-6,74	79
2	21,25	-6,34	41
3	15,38	-8,85	423
4	30,61	-6,73	130
5	24,94	-13,68	95
6	28,60	-11,19	99
7	28,60	-11,19	36
8	16,47	-6,44	43
9	27,12	-8,20	111
10	15,79	7,53	34
11	16,59	-11,57	86
12	25,85	4,09	124

Fig. 6 Cluster summary with cluster centroid calculated

Source data included 264127 rows describing customers, out of which 174297 were assigned to clusters by HDBSCAN algorithm. The remaining customers (89830) were assigned to cluster -1, which means that insufficient density was found in the area they were located (no pattern was detected). 2046 small but very consistent clusters were discovered. Such quantity of clusters is too high to effectively address managers' needs, hence the reason why the aim of the next step will be to aggregate small clusters into customer communities.

#### IV. DISCOVERY OF CUSTOMER COMMUNITIES - LABEL PROPAGATION

In [27], communities are defined as "*groups of vertices within which connections are dense, but between which connections are sparser*". According to [28], such communities can be considered as fairly independent spaces of a graph, sharing common properties and/or playing similar roles within it.

In our study, communities are groups of customer clusters whose elements share common properties and allow managers to apply the same measures to them or to identify strong similarities between groups in the same community.

Label Propagation Algorithm has been proposed by Raghavan et al. [14] for detecting communities in networks represented by graphs. The algorithm, due to its linear time complexity of  $O(m)$  for each iteration, simplicity, and ease of implementation, is commonly used to identify communities in large-scale real-world networks, such as social media.

An advantage of the algorithm is that it does not require prior information about number of communities or their cardinalities to run; neither does it require any parameterization. The number of iterations to convergence is barely dependent on the graph size, but it grows very slowly.

In [29] the LPA has been compared with other clustering algorithms: Louvain algorithm [30], Smart Local Moving (SLM) [31] and Infomap algorithm [32]. Results of that experiment favors LPA to be used with large scale data as it outperforms other algorithms for well-defined clusters.

These characteristics of the LPA method was the main reason for choosing it for detecting communities of customers and proposing a new method combining multiple methods: HDBSCAN creating numerous clusters, UMAP reducing dimensions, k-NN forming graph, and LPA finding communities.

The main idea behind LPA is to propagate labels throughout the graph from a node to its neighbor nodes. As a result, the groups of nodes sharing the same label and whose nodes have more neighbors than nodes in other groups make communities.

The algorithm consists of five steps [33]:

1. Initialize the labels at all nodes in the network. For a given node  $x$ ,  $c_x(0) = x$
2. Set  $t=1$

3. Arrange the nodes in the network in a random order and set it to  $x$
4. For each  $x \in X$  chosen in that specific order, let
 
$$c_x(t) = f(c_{x_1}(t), \dots, c_{x_m}(t), c_{x_{i(m+1)}}(t-1), \dots, c_{x_k}(t-1))$$
 where  $f$  returns the label occurring with the highest frequency among neighbors. Select a label at random if there are multiple highest frequency labels.
5. If every node has a label that the maximum number of their neighbors has, then stop the algorithm. Else, set  $t=t+1$  and go to (3).

Label propagation works as follows: at the beginning all clusters make own communities, by assigning unique labels to every cluster, then the following steps are being executed in a loop. In every iteration all clusters are processed in a random order and the labels are updated to one that occurs with the highest frequency among the direct neighbours. If the label cannot be chosen as there are multiple labels occurring with the same frequency, then one of them should be chosen randomly. If all clusters are processed in this iteration, stop condition is checked: all clusters should be labelled with the one that majority of adjacent clusters have and if the condition is met, the algorithm ends. Otherwise, the iteration is repeated until convergence defined as the stop condition is reached. In this way, labels will propagate across the graph, replacing other labels and eventually some labels will disappear, and others will dominate.

It is important to note that Label Propagation Algorithm operates on graphs, hence the input data must be converted into a graph. In our experiment, it was necessary to performed on "fashion vs. discount" case study a dimension reduction with UMAP, grouping customers with similar properties into clusters and determining centroids of each cluster accordingly.

In order to create a graph, the "k-Nearest Neighbors algorithm" (known as k-NN) was used that is one of the simplest, but perfectly fitting into the experiment context, and as a non-parametric method it is commonly used for classification and regression. For classification, the centroids with Euclidean distance between them are used and transformed into the normalized distances for all nodes while filtering out all that above a given threshold. More details about the results of LPA and k-NN on real marketing data will be given in section 5.

A graph was created, where the nodes represent clusters generated by HDBSCAN and edges are weighted links connecting clusters, determined by applying k-NN algorithm from the previous step and representing normalized distances between clusters. Centroids defined while executing UMAP method were crucial in the creation of a proper graph for LPA method.

Finally, using the data from the previous steps, a large graph was created, consisting of 2046 clusters (vertices) and

15364 links as represented in Fig. 7. The graph is shown only for demonstrational purposes, where lengths of links do not illustrate the real distances between clusters (weights), nonetheless cluster proximities have been preserved. Multiple dense cluster groups can be noticed. These are candidates to form communities and in compliance with the definition of a community, they have many connections within the group and few to clusters outside the group. More detailed information about the graph structure will be given in the next section.

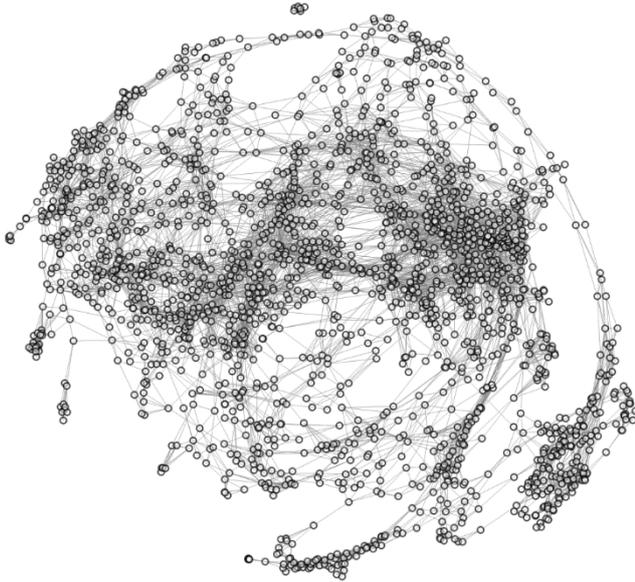


Fig. 7 Visualization of the graph representing communities and connections between them

Densely connected groups reach a common label quickly. When many such dense groups are created throughout the network, they continue to expand outwards until it becomes impossible to do so. Randomization of the order the clusters are processed has consequences: it may not deliver a unique solution, or the final solution may not be found at all (due to fluctuations in label assignment, adjacent clusters can interchange their labels in every iteration, preventing the convergence criteria from being achieved).

In our tests all partitions found were similar to each other, though.

Finally, the graph looks as in Fig. 8. The densest groups of clusters have been marked in color on the graph. They form communities characterized by common attributes.

## V. INTERPRETATION OF THE EXPERIMENTAL RESULTS

In retail businesses, managers would want to know about the customers in order to efficiently tailor offers for selected groups, and to increase efficiency and customer satisfaction, which in turn increases business profitability. On the other hand, there might be a group of customers that may abuse the system, searching for system weaknesses and resulting in a loss or very small benefit for the retailer. The promise of this experiment was to find clusters of most profitable

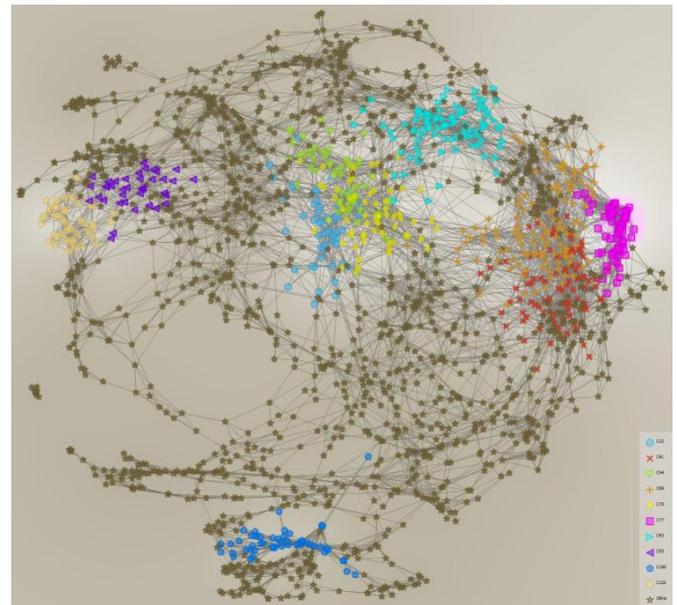


Fig. 8 Graph of clusters with highlighted communities

customers what has been already stated in previous part, and to provide marketing manager or analyst with appropriate knowledge about customer behavior respecting the value of products, discount, and “age” of the product.

In the experiment above Label Propagation Algorithm has been applied on clusters of customers having similar purchase characteristics and identified groups (communities) of similar clusters.

Let us analyze two groups: C77 and C122 among all groups identified by LPA in the experiment, as visualized below (Fig. 9 and Fig. 10).

First group C77, marked in pink color, consists of customers buying goods present in the shop for several months, but always with a price discount. The second group C122, marked in orange, seems to be very similar to C77, but it represents customers buying goods with the highest price discounts. For these two groups the measures applied by managers should be different. For the first group it could be running a marketing campaign in order to increase the average price of the order, while the second group can be used to address seasonal sale campaigns at the late stage

If a manager is interested in clusters of customers with the highest order values (as potentially most beneficial customers), they can filter clusters using the order-value property. Selected clusters of customers who buy goods valued at more than 350 PLN are presented in Fig. 11. There are 133 groups (communities) created by LPA out of 2052 clusters generated by HDBSCAN. On the left-hand side of visualization one dot represents one HDBSCAN cluster and the color of the dot represents LPA cluster (after community detection by LPA).

If a manager is interested in some specific clusters, they can observe the distribution of each feature in clusters. If one takes into consideration clusters C2 and C7 presented in Fig. 11, one can observe that these are customers not looking



Fig. 9 Clusters of C122 group forming a community



Fig. 10 Clusters of C77 group forming a community

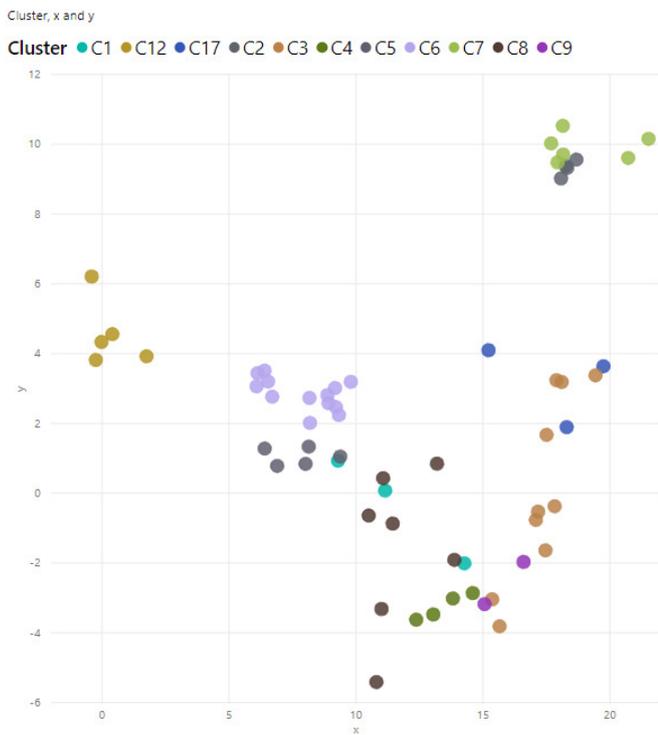
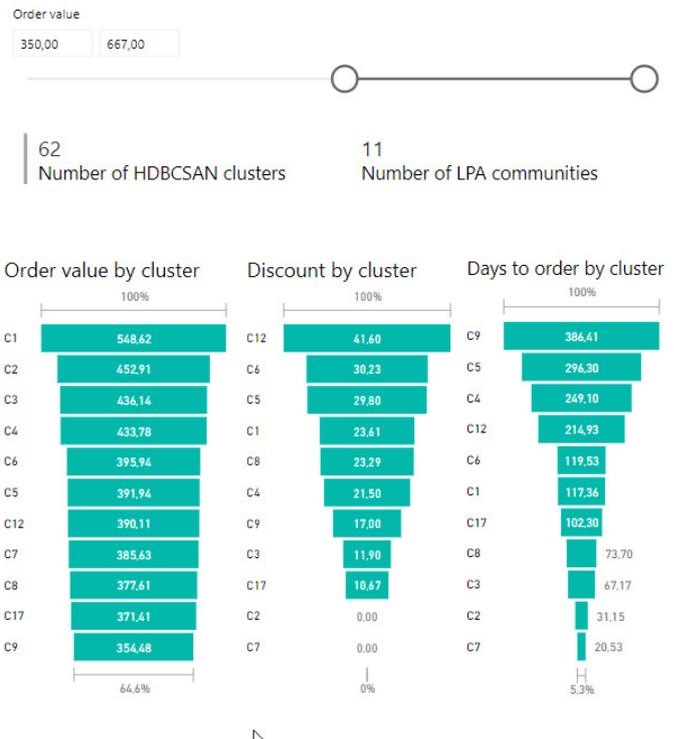


Fig. 11 Graphical representation of analyzed clusters meeting manager's criteria



for discounts (they buy with 0% discount), they buy new products (launched accordingly 31 and 20 days before purchase). The difference between those clusters is in the average order value (respectively 452 and 385 PLN). Having such knowledge, the recommendation system makes it possible to tailor the offer in order to meet customer's expectations.

For seven selected clusters meeting the assumed criteria (Fig. 12), the scatter within the cluster was calculated using Davies-Bouldin index in order to compare results with k-means result. In this experiment, the measure of scatter was between 14.44 and 46.27. Average value for selected 7 clusters is 34.13 which in comparison with the value of k-means 80.47 constitutes a significant improvement.



Fig. 12 Interpretation of selected clusters

## VI. CONCLUSIONS AND FUTURE RESEARCH

The primary objective of the presented research was to develop a method to discover meaningful customer communities, using data mining techniques and tools. The outcome of this work is a new multi-level hybrid method for community discovery, implemented and validated on the experimental platform Upsaily. The research methodology is composed of six, closely integrated steps. Firstly, relevant information about customers is extracted from large marketing databases and partitioned by the clustering algorithm (HDBSCAN). Secondly, the space dimensions are reduced into two dimensions using the Uniform Manifold Approximation and Projection (UMAP) method. Thirdly, the centroids are computed for each cluster. The graph is generated in step four using the k-NN algorithm. To discover customer communities the Label Propagation Algorithm (LPA) is applied. The final step, most important for decision makers, concerns marketing interpretation of discovered customer communities. These experiments demonstrated that the “customer communities discovery” compared against “segmentation with k-means algorithm”, gave much more precise identification of group of customers and allows better understanding of clusters by managers and data analysts.

The multi-level clustering approach described in this paper has shown its advantage over single method clustering. Numerous small clusters were turned into communities sharing common properties. Specifically, running HDBSCAN alone against the data describing customer’s purchases resulted in a high number (2046) of dense, but small clusters, making it infeasible to predict customer’s needs or address tailored offerings. It is important to mention that using the simplest PCA method for dimensions reduction did not meet expectation. The clusters did not form homogenous communities, which in turn could not provide managers with reliable tools to support decision making processes. When the PCA method was replaced by the UMAP method, the clustering results met expectations and made it possible to calculate meaningful centroids for each cluster. Afterwards, the label propagation method was applied, making it possible to

determine customer communities, grouping them based on business needs.

The Upsaily platform used in the experiment, allows for parameterization of the multi-level approach to clustering, described in the paper, by defining the features used for clustering, business-oriented cluster identification, defining data range or specifying the size of expected clusters. The advantage of this customized approach is that it can be widely applied to any type/category of customers and it allows for performing advanced analytics on the business data.

The results obtained so far on real marketing data are very encouraging, in addition they have been positively validated by managers of internet shops. However, many algorithmic and business-oriented issues remain to be extended and tuned. For instance, a desirable extension of the approach will be to refine a method of feature construction describing a customer profile. An interesting future improvement will be on the implementation of collective and cooperative clustering with built-in business-oriented quality measures. One, but not the last, ambitious work will be focused on the dynamics and evolution of customer communities.

## REFERENCES

- [1] Barber M. J. (2007). Modularity and community detection in bipartite networks, *Physical Review E*, 76(6):066102, DOI:10.1103/PhysRevE.76.066102
- [2] Codaasco G., Gargano L. (2011). Label propagation algorithm: A semi-synchronous approach, *Internat. Journal of Social Network Mining*, 1(1):, pp.3-26, DOI:1504/IJNSM.2012.045103
- [3] Gregory S. (2010). Finding overlapping communities in networks by label propagation, *New J. Phys.*, 12, 103018, DOI:10.1088/1367-2630/12/10/103018
- [4] Han J., Li W., Su Z, Zhao L. and Deng W. (2016). Community detection by label propagation with compression of flow, e-print arXiv:161202463v1, DOI:10.1140/epjb/e2016-70264-6
- [5] Liu W., Jiang X., Pellegrini M., Wang X. (2016). Discovering communities in complex networks by edge label propagation, *Scientific Reports* 6, DOI:10.1038/srep22470
- [6] Rossetti G., Cazabet R. (2017). Community Discovery in Dynamic Networks: A Survey, arXiv:1707.03186, DOI:10.1145/3172867
- [7] Wu Z.H. et al. (2012). Balanced multi-label propagation for overlapping community detection in social networks, *Journal of Comp. Sc. And technology*, 27(3), pp. 468-479, DOI:10.1007/s11390-012-1236-x
- [8] Subelj L., Bajec M. (2014). Group detection in complex networks: An algorithm and comparison of the state of the art, *Physica A: statistical Mechanics and its Applications*, 397, pp. 144-156, DOI:10.1016/j.physa.2013.12.003
- [9] Gordon S., Linoff M., Berry J.A. (2011). *Data Mining Techniques for Marketing, Sales, and Customer Relationship*, Wiley, ISBN:978-0470650936
- [10] Aggarwal C.C., Reddy C.K. (2013). *Data Clustering: Algorithms and Applications*, Chapman & Hall / CRC, ISBN:978-1466558212
- [11] Gan G., Ma C., Wu J. (2007). *Data Clustering: Theory, Algorithms, and Applications*, SIAM Series, DOI:10.1137/1.9780898718348
- [12] Pondel M., Korczak J. (2018). Recommendations Based on Collective Intelligence—Case of Customer Segmentation. In *Information*

- Technology for Management: Emerging Research and Applications (pp. 73-92). Springer, Cham, DOI:10.1007/978-3-030-15154-6\_5
- [13] Witten I.H. et al. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, ISBN:978-0128042915
- [14] Raghavan U.N., Albert R., Kumara S. (2007). Near linear time algorithm to detect community structures in large-scale networks, *Phys. Rev. E*, 76,036106, DOI:10.1103/PhysRevE.76.036106
- [15] Rosvall M., Bergstrom C. T. (2007). An information-theoretic framework for resolving community structure in complex networks, *Proc. Natl. Acad. Sci.*, 104, pp. 7327-73-31, DOI: 10.1073/pnas.0611034104
- [16] Xie J.R., Szymanski B.K. (2014). LabelRank: a stabilized label propagation algorithm for community detection in networks, *Proc IEEE, Network Science Workshop*, pp. 386-399, DOI: 10.1109/NSW.2013.6609210
- [17] Applebaum W. (1951). Studying customer behavior in retail stores. *Journal of marketing*, 16(2), 172-178, DOI: 10.2307/1247625
- [18] Clover V.T. (1950). Relative importance of impulse-buying in retail stores. *Journal of marketing*, 15(1), 66-70, DOI: 10.1177/002224295001500110
- [19] See-To E., Ngai E. (2019). An empirical study of payment technologies, the psychology of consumption, and spending behavior in a retailing context. *Information & Management*, 56(3), 329-342, DOI: 10.1016/j.im.2018.07.007
- [20] Rustagi A. (2011). A Near Real-Time Personalization for eCommerce Platform. In *International Workshop on Business Intelligence for the Real-Time Enterprise* (pp. 109-117). Springer, Berlin, Heidelberg, DOI: 10.1007/978-3-642-33500-6\_8
- [21] Kaptein M., Parvinen P. (2015). Advancing e-commerce personalization: Process framework and case study. *International Journal of Electronic Commerce*, 19(3), 7-33, DOI:10.1080/10864415.2015.1000216
- [22] Campello R.J., Moulavi D., Sander J. (2013, April). Density-based clustering based on hierarchical density estimates. In *Pacific-Asia conference on knowledge discovery and data mining* (pp. 160-172). Springer, Berlin, Heidelberg, DOI:10.1007/978-3-642-37456-2\_14
- [23] Pondel M., Korczak J. (2017). A view on the methodology of analysis and exploration of marketing data. In: *Federated Conference on co-algorithm to detect community structure in large-scale networks*, *Phys.Rev. E*. 760360106 *Computer Science and Information Systems (FedCSIS)*, IEEE, pp. 1135-1143, DOI:10.15439/2017F442
- [24] Schubert E., Sander J., Ester M., Kriegel H.P., Xu X. (2017). DBSCAN revisited, revisited: why and how you should (still) use DBSCAN. *ACM Transactions on Database Systems (TODS)*, 42(3), 19, DOI:10.1145/3068335
- [25] Ester M., Kriegel H.P., Sander J., Xu X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd* (Vol. 96, No. 34, pp. 226-231), ISBN:1-57735-004-9
- [26] McInnes L., Healy J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. Preprint arXiv:1802.03426
- [27] Newman M.E.J. (2004). Detecting community structure in networks. *Eur. Phys. J. B* 38(2), 321-330, DOI:10.1140/epjb/e2004-00124-y
- [28] Fortunato S. (2004). Community detection in graphs. Preprint arXiv:0906.0612, DOI:10.1016/j.physrep.2009.11.002
- [29] Emmons S., Kobourov S., Gallant M., Börner K. (2016). Analysis of Network Clustering Algorithms and Cluster Quality Metrics at Scale. *PLOS ONE* 11(7): e0159161. DOI:10.1371/journal.pone.0159161
- [30] Blondel V.D., Guillaume J.L., Lambiotte R., Lefebvre E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*. 2008(10):P10008. DOI: 10.1088/1742-5468/2008/10/P10008
- [31] Waltman L., Eck N.J. (2013). A smart local moving algorithm for large-scale modularity-based community detection. *The European Physical Journal B*. 86(11):1-14. DOI:10.1140/epjb/e2013-40829-0
- [32] Rosvall M., Bergstrom C.T. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*. 105(4):1118-1123. DOI:10.1073/pnas.0706851105
- [33] Zhu X., Ghahramani Z. (2002). Learning from labeled and unlabeled data with label propagation (p. 1). Technical Report CMU-CALD-02-107, Carnegie Mellon University



# ICT Usage in Industrial Symbiosis: Problem Identification and Study Design

Linda Kosmol

Technische Universität Dresden  
Chair of Business Informatics, esp. Systems  
Engineering  
01062 Dresden, Germany  
linda.kosmol@tu-dresden.de

Christian Leyh

Technische Universität Dresden  
Chair of Business Informatics, esp. IS in  
Manufacturing and Commerce  
01062 Dresden, Germany  
christian.leyh@tu-dresden.de

□

**Abstract**—Industrial symbiosis is a favored approach to balancing industry’s economic growth and its environmental impact on a regional scale. Although the scientific literature reports a multitude of examples of industrial symbiosis around the world, this approach and its related concepts are not considered to be widespread in practice, due to various barriers. Information and management barriers are seen as significant obstacles to industrial symbiosis; however, they have not been adequately investigated yet. Empirical research capturing the perception of industrial actors is lacking. This applies especially to information and communication technology designed to reduce informational barriers. Therefore, in this research-in-progress paper, we will first examine these aspects by discussing related publications. In the second step, we develop a study design involving an online questionnaire to examine the extent of managerial and informational barriers that prevent industrial symbiosis as well as the perception of corresponding technological support.

## I. INTRODUCTION

The term "sustainability" has become omnipresent and a central issue in discussions on industrial development. In this regard, one approach to sustainable industrial development that is attracting more and more research attention is industrial symbiosis. Industrial symbiosis aims to balance industrial activities and their impact on the environment. It involves regional, cross-sectoral, and cross-company cooperation to increase resource efficiency and to ecologically and economically benefit the parties involved [1], [2]. As a result, this approach can create so-called ‘industrial ecosystems’ or ‘eco-industrial parks’ from conventional industrial systems or parks, in which local industries adopt symbiotic behavior and commit to sustainable development policies [3], [4].

Nevertheless, reports show that industrial symbiosis is not widely implemented in practice, due to technical/physical, financial, economic, regulatory, social, informational, and managerial barriers [5], [6]. According to [7], less than 0.1% of 26 million active enterprises in Europe are engaged in industrial symbiosis. In particular, managerial (e.g., limited commitment) and informational (e.g., lack of information/knowledge sharing) barriers are regarded as

significant obstacles to industrial symbiosis [5]–[10], since they must be overcome for symbiotic opportunities to be identified. If these barriers are not addressed, subsequent processes—such as feasibility studies of identified opportunities—cannot be carried out and other related barriers (such as technical or financial issues) cannot be identified.

Different examples of industrial symbiosis around the world have been described in the scientific literature [11], [12]. These examples contribute to sustainable development in economic (e.g., reduced waste disposal and input costs), environmental (e.g., reduction waste production and resource use), and social (e.g., community awareness) terms. However, at this point, industrial symbiosis research needs more empirical and quantitative studies, which are currently lacking (as stated by [13]). In recent years, some surveys on industrial symbiosis activities and barriers have been conducted, both to examine the interest in and maturity of industrial symbiosis in a specific region and to capture the perception of barriers to adopting symbiotic behavior (see Table I). The surveys usually address companies (general or environmental managers) or policymakers but rarely the managers of industrial parks, despite the wide recognition of their potential role as facilitators [14], [15].

TABLE I.  
INDUSTRIAL SYMBIOSIS SURVEYS ON INDUSTRIAL SYMBIOSIS

Year	Article	Region	Focus <sup>1</sup>	Sample <sup>2</sup>
2016	[16], [17]	Philippines	Barrier interdependencies (10)	10
2017	[18]	Brazil	Social barriers (4)	29
	[14]	Europe	Symbiosis activity, barriers (5)	92
	[15]	Europe	Symbiosis activity, barriers (10)	n/a
2018	[19]	Slovenia	Symbiosis activity (-)	50
	[20]	Sweden	Symbiosis activity, maturity, barriers (4)	20 (50)
	[21]	Spain	Symbiosis activity, barriers (9)	95
2019	[22]	Europe	Symbiosis activity, impacts, barriers (12)	22 (25)

□ This work was not supported by any organization.

<sup>1</sup> No. in brackets = No. of barrier categories

<sup>2</sup> No. in brackets = No. of follow-up interviews

Financial/economic and regulatory barriers are considered the largest barriers by many study participants. In terms of informational and managerial barriers, views are mixed, with large variation in perceived relevance. These differences can be caused by region and context, but also by the different barrier categorizations proposed in the different studies. Informational, managerial, and social aspects are not clearly separated with regard to how barrier categories are assigned. Informational barriers in particular are often linked with social factors such as trust, cooperation, and community. Although a cause-effect relationship may exist (according to [14] and [15], managerial barriers are a causal factor and informational barriers are an effect factor), the present inconsistency in barrier categorization and the lack of a uniform allocation of the underlying aspects make it difficult to compare the studies, reducing their clarity and validity.

Only [19] and [21] considered informational barriers separate from social aspects, albeit exclusively with reference to information systems. The survey of [21] reveals that some companies consider inadequate information management systems to be a barrier. The survey of [19] shows that 49 out of 50 respondents would use an online platform to search for potential partners if one was available. However, in both studies, no further information is given regarding the information system/online platform (information types, functionality, access structure, users, etc.). Therefore, the extent of informational barriers and their cause remain unknown.

A similar lack of empirical foundation can be seen in the efforts to support industrial symbiosis with information and communication technology (ICT), as a way to mitigate informational issues. These tools appear to be primarily research-driven, and the extent to which they are known, used, and judged useful by companies remains unclear [8], [23]. Therefore, we decided to set up a long-term research project to address this research gap. In our opinion, research must further explore the readiness of companies and managers in industrial parks to practice industrial symbiosis, as well as the opinions on informational issues and ICT support. An investigation into informational aspects regarding information availability, confidentiality, and relevance, as well as on the perception of ICT support, will especially contribute to understanding and uncovering gaps between research efforts and practice. Therefore, the aim of this research-in-progress paper—the first step of our research project—is to identify and clarify the necessary aspects of informational and managerial barriers and of ICT support for industrial symbiosis, in terms of problem identification. In addition, we will present our initial study design (an online questionnaire) and outline the next steps of our ongoing research.

This paper is structured as follows. Section II presents the theoretical background on industrial symbiosis and its barriers in terms of problem identification. Then, Section III describes our study design. Finally, Section IV concludes the paper with a discussion and future steps.

## II. THEORETICAL BACKGROUND AND PROBLEM IDENTIFICATION

### A. Business Models in Industrial Symbiosis

Originally, the term ‘industrial symbiosis’ covered the physical exchange of material, energy, water, and by-products between geographically close companies in order to achieve economic and environmental advantages [1]. Today, the term encompasses all business models of inter-firm exchange or sharing of under-utilized resources like material, energy, logistics, capacities, space, expertise, and knowledge [2]. The business models (synergies) are therefore either exchange-based or sharing-based and are commonly divided into three categories [24], [25]:

- By-product exchange and reuse
- Utility and infrastructure sharing
- Service sharing

*By-product exchange* refers to one company’s residual outputs (e.g., waste and by-products) being used as another company’s inputs (e.g., water, material, waste heat). Here, the principle of circular economy is followed. *Utility and infrastructure sharing* refers to joint use and/or operation of technical infrastructure and decentralized plants, such as a combined heat and power plant, water treatment plant, district heating grid, etc. *Service sharing* refers to cross-company management or joint provision of common services, (e.g., joint disposal/procurement, logistics and warehousing, staff training, knowledge exchange).

Depending on the business model and the role of the company (e.g., supplier or consumer), various economic and ecological advantages can arise for the parties involved and for the corresponding region: reduced resource consumption and waste generation, eco-innovation, revenues from residues and by-products, less raw material and disposal costs, development of new business and market opportunities, etc. [26].

The development mechanisms of industrial symbiosis can be divided into three categories—*self-organized*, *facilitated*, and *planned/designed*—with the degree of the involvement of coordinating/mediating third parties (e.g., research institutes, governmental agencies, park management) increasing [27]. Intermediaries are regarded as vital to supporting contact initiation, to collecting necessary information and knowledge, and to facilitating their exchange between companies [28]. Park managers are considered to be the best candidates to provide this social and informational infrastructure to the companies in an industrial park [5].

### B. Informational and Managerial Barriers to Industrial Symbiosis

#### 1) Informational Barriers

The availability of information and knowledge, as well as the willingness to share them with others, is essential to identifying and evaluating synergy opportunities for industrial symbiosis. For example, in the case of a potential

by-product exchange, this would include information on the incoming and outgoing resource flows of companies, as well as knowledge of relevant compatibility criteria and technical expertise to implement synergies. However, lack of trust, confidentiality issues, and motivational issues may lead to an unwillingness to share necessary information and knowledge [5], [28], [29]. Lack of internal information, lack of contacts and relationships with whom to share information and knowledge, communication issues, and difficulty to share knowledge limit available information and knowledge sources [6], [8], [30]. Confidentiality issues in particular have not been investigated in the context of industrial symbiosis.

Since we want to determine to what extent informational barriers exist, we want to keep them as separate as possible from social aspects, such as trust and relationships. Therefore, we consider the following issues to be informational barriers to industrial symbiosis:

- Unawareness of principles and benefits of industrial symbiosis
- Lack of available information and knowledge
- Unwillingness to and difficulty of sharing information and knowledge
- Non-transparency and inefficiency of the information and knowledge exchange
- Lack of information-sharing mechanisms and infrastructure

ICT support is considered to be promising in alleviating informational barriers and providing a space for interaction and exchange between companies [9], [10]. Generally, these tools (developed or conceptualized) are online repositories (e.g., ISData) and platforms (e.g., eSymbiosis) that provide various functions for disseminating and sharing information and knowledge as well as for facilitating byproduct exchanges via waste market functions and automatic matching engines.

However, industrial symbiosis ICT tools face various barriers [31], [32]. Currently, these ICT tools are not provided with enough data and information for them to be used effectively. This may be caused by lack of willingness to use the tools, confidentiality issues, manual effort required for data entry, lack of knowledge of the existence of such tools, and access restrictions—thus leading to a low number of potential users. Social networking approaches [33], [34] have increasingly addressed criticisms of early tools for not taking sufficient account of the social context [8]. In addition, many tools limit their functionality to the early stages of industrial symbiosis (synergy identification and assessment) [8], [10]. Moreover, many tools are not easily/publicly accessible, not operational, or still in the concept or development stages [8], [10]. Hardly any statements can be found on the operational tools as to the context in and extent to which they are used, how and by whom, and which specific functions they provide to (potential) users. Therefore, it is difficult to assess how useful current ICT support is to industrial symbiosis.

## 2) Managerial Barriers

In order to gather and share information and knowledge, there must be a willingness to commit to sustainable business models; to participate in workshops; and to provide time, personnel, and (likely) financial resources. Synergy identification and implementation do not only deal with resources, they also require resources. Without commitment to incorporating the concept of industrial symbiosis into a holistic strategy and business processes of participating companies, the discovered potential synergies may remain unused [8]. Since the potential benefits are unknown and difficult to predict at first, and since coordination and the exchange of information and knowledge are time-consuming, this commitment and willingness must continue beyond initial meetings and workshops. Therefore, we consider the following (organizational) aspects as managerial barriers:

- Lack of commitment to sustainable business and to the community/network
- Lack of management support
- Unwillingness to collaborate and communicate

The attitude of the company and park management not only influences the extent of informational barriers but can also determine how—and if—ICT tools are used for industrial symbiosis.

## III. RESEARCH APPROACH

### A. Overall Research Approach

In our research project, we pursue the design of an ICT tool (IT artifact) to mitigate and overcome informational and managerial issues in industrial symbiosis. Therefore, we follow the design science paradigm [33]. The steps for design science in information systems research are shown in Fig. 1.

Currently, we are in the first stages of our long-term research project. We initiated our research by identifying the problem (managerial and informational barriers) and solution space (ICT) in industrial symbiosis through a comprehensive literature review and discussions with industrial partners. As described in Section II, managerial and informational barriers still persist, and ICT solutions encounter a number of problems. Furthermore, design guidelines and best practices for ICT solutions that enable/support industrial symbiosis are lacking [31].

Like other researchers, we believe that ICT tools, particularly digital platforms, can contribute to overcoming managerial and informational barriers, but only if the design is tailored to the needs, circumstances, and restrictions of the intended users. Capturing these aspects requires capturing the general attitude of management towards industrial symbiosis along with the associated exchange of information and knowledge and the corresponding ICT. To this end, we will conduct a survey with the relevant industrial players in industrial symbiosis in industrial parks. This problem-centered approach underlines the relevance of the topic and clarifies the problems addressed in our research project.

Based on the answers and the information obtained in subsequent interviews, we aim to deduce where an ICT tool could and should be applied, and what it could achieve in terms of enabling industrial symbiosis. By conducting empirically-grounded problem identification and aligning it with the existing solution space, we aim to reconfirm our

research gap—thus thoroughly justifying our project—and to ensure that the ICT tool is designed in such a way that it can be embedded in existing strategies and processes to address this important issue [23].

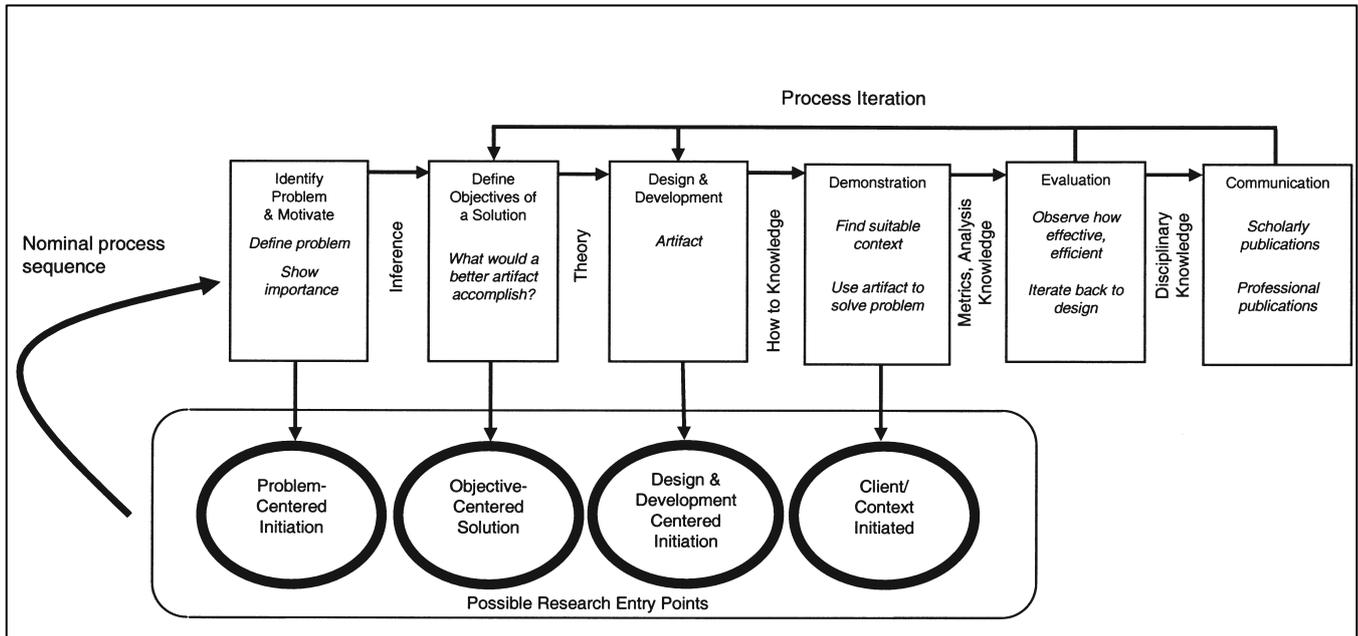


Fig. 1 Design Science Research Methodology [33]

### B. Proposal of Study Design – Online Questionnaire

To address the identified problems/barriers and to investigate them in more detail, thus filling this gap in the literature, we set up a study design based on an online questionnaire. The aim of this approach is to reach a large number of companies and to gather various opinions and perceptions of managerial, informational, and ICT-related aspects in the context of industrial symbiosis. The questionnaire will be provided in German for companies in Germany, Austria, and Switzerland and in English for other European companies. It will primarily be sent to companies in industrial parks and to park managers, as these companies and park management are predestined for industrial symbiosis. We aim to gain insights into the (estimated) willingness and ability of companies to cooperate in industrial symbiosis and into the willingness and ability of park management to act as a facilitator. The questionnaire is composed of four sections:

#### 1) General Data – Participant Characterization

In order to accurately analyze the answers of the participants, the participants themselves must be sufficiently characterized. Therefore, the general data section includes, for example, the following information:

- Size of company/industrial park (number of employees/companies), which indicates the human resources available and the number of potential synergy partners
- Length of stay at site (in years), which may influence the number of established contacts and (business) relationships
- Certification in energy (ISO 50001) and/or environmental management (ISO 14001/EMAS) and acknowledgement of the importance of energy and material consumption and waste, which indicates the relevancy of sustainability issues.

The characteristics of enterprises and the comparisons between them may reveal fundamental differences in readiness as well as in business and information attitudes towards industrial symbiosis, from which different conclusions can be drawn (e.g., requirement profiles).

#### 2) Managerial Aspects – Current Practice, Readiness, and Potentials

The extent of management-related barriers is reflected in the current practices and the readiness to adopt industrial symbiosis practices and business models. Therefore, the first block will question the current practice, including the type of the business model and the role of the company. This block

also examines awareness of the concept of industrial symbiosis regardless of specific terms.

Subsequently, the survey participants will be asked to assess the readiness, interest, and potential opportunities of the company/companies to practice industrial symbiosis. Questions related to readiness indicate a company’s ability to collaborate at the company level, while questions related to potential assess the company’s ability to collaborate at the network level (by questioning the perceptions of other companies in the industrial park). To measure readiness and potential, we use the proposed readiness areas of [34]. These areas involve the business models of industrial symbiosis (e.g., readiness for by-product exchange) and the company’s strategic orientation towards industrial symbiosis (e.g., readiness to pursue common goals or to provide time and personnel for industrial symbiosis activities). The answer options will use a 5-point Likert scale indicating low (1) to high (5) readiness/potential.

The questions regarding potential also help indicate the extent to which companies have information/knowledge of other companies at their location/industrial park. A screenshot of example questions in this section is provided in the Appendix (see Fig. 5).

3) Informational Aspects – Availability and Sharing

The extent of information-related barriers is reflected in concept awareness, internal and public availability of information and knowledge, and willingness to disclose the latter.

First, the company’s policies and practices in terms of internal and external exchange of knowledge are examined. ‘Policy’ refers to management support for facilitating knowledge sharing, while ‘practice’ refers to existing formal and informal communication channels and methods of knowledge sharing. Companies could benefit from improving or qualifying existing channels and practices, instead of developing and imposing new ones. An example question on communication channels is given in Fig. 2, and a screenshot of this question is provided in the Appendix (see Fig. 6).

<b>Which communication channels do you mainly use in your company to exchange experiences and knowledge with colleagues and partners?</b>					
	<b>Within your company</b>				
	Never	Rarely	Sometimes	Often	Always
	1	2	3	4	5
Face-to-face communication	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Virtual face-to-face communication	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
E-mail	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Intranet	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Expert systems	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Fig. 2 Example Survey Question – Knowledge Sharing

Second, the availability of needed information is investigated. To this end, we will ask whether certain information about inputs and outputs is known within the company, whether it is already published/publicly available (e.g., in environmental reports), and whether it is generally subject to confidentiality. The types of information to be addressed are listed in Table II. This information is typically used to identify synergies, but at a high level of detail [35]. In order to examine the willingness to share information, we will include questions on the relevance of the types of information, as well as questions on the disclosure of information and its level of detail. An example question concerning information confidentiality is given in Fig. 3, and a screenshot of this question is provided in the Appendix (see Fig. 7).

TABLE II. NECESSARY INFORMATION FOR INDUSTRIAL SYMBIOSIS

Information type	Examples/Level of detail
Resource type	Material, energy, water, EWC classification
Resource quantity	Average per year/per month/per day/per hour
Supply pattern	Constant/fluctuating, maximum/minimum, lot size
Resource property	Components/ingredients, pollution, temperature
Resource source	Plant type (e.g., processing/production plant), utilization (e.g., material input, drying, air conditioning, process heat), specific plant (e.g., industrial furnace)
Availability period	All year/seasonal, month details (e.g., April-August), date specification (e.g., 01.04.19-04.12.20), shift system (e.g., Mo-Fr 5:30 to 22:30)
Supplier/customer	Type, name
Price/cost	Total per year/per unit, upper/lower price limit

<b>Are the following types of information classified as confidential in your company and therefore subject to disclosure restrictions to other companies (e.g., non-disclosure agreements)?</b>			
	<b>Raw materials</b>		
	Yes	Uncertain	No
Resource type	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Resource quantity	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Supply pattern	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Resource properties	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Availability period	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Supplier/customer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Price/cost	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Fig. 3 Example Survey Question – Information Sharing

Another important aspect related to information sharing is the type of information exchange:

- Direct: Other interested companies in the industrial park can receive or view information.
- Indirect: An intermediary receives the information, processes and analyzes it, and passes on the results of the analysis (e.g., that a synergy potential is presumed or not) to other companies, without the latter having access to the original information

Companies' preference for one type or the other can provide important insight into a suitable mechanism for information exchange in an industrial park and for the design of appropriate ICT support.

#### 4) ICT Support – Awareness and Design

In this section of the questionnaire, the awareness of ICT tools and the perception their usefulness are addressed on the basis of provided functions.

Based on [9], different types of ICT tools (e.g., online waste market, synergy identification system) will be presented with examples, in order to assess participants' awareness of them. Subsequently, selected functions of these tools (e.g., exchange market, matching engine, social applications) will be presented, in order to examine whether these functions are considered useful and/or would be used. At this point, a question is included inquiring whether a direct or indirect exchange of information is preferred. The questionnaire for both the companies in industrial parks and the park management should include a question of who should provide and operate such a tool (see Fig. 4).

<p><b>Who do you think should operate/provide such systems or platforms?</b></p> <p><input type="checkbox"/> Park management</p> <p><input type="checkbox"/> Focal company</p> <p><input type="checkbox"/> Third Party</p> <p><input type="checkbox"/> Uncertain</p>
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Fig. 4 Example Survey Question – ICT

Since the questionnaire will be tailored to two different target groups (companies in industrial parks and park management), some questions will vary slightly according to the role of the participants, and some will only be accessible to one target group or another. While the companies in industrial parks will assess their own readiness for industrial symbiosis and their information level, the park managers will assess the readiness and the information levels of the companies located in their park, as well as their willingness to act as facilitator in coordinating industrial symbiosis activities or providing services.

#### IV. DISCUSSION AND FUTURE STEPS

In the existing industrial symbiosis research, managerial and informational issues and associated ICT support are not sufficiently addressed. In particular, empirical research capturing industrial actors' perceptions of these barriers and ICT is lacking. In order to develop an appropriate ICT tool to overcome the barriers (the aim of our long-term research), it is necessary to identify the specific underlying problems, needs, and resistances/aversions of the potential users.

Our first discussions with industrial park members and managers have shown that a general interest and a willingness to cooperate are present, but the human, time, and financial resources necessary to pursue and adequately implement the

concept are rarely made available in the companies. These conversations also revealed that the awareness of the concept of industrial symbiosis is not perceived as an issue; however, the exchange of information was always seen as problematic and in need of improvement. Furthermore, none of the discussion partners knew of existing ICT tools, but they imagined that using them would be beneficial. These findings confirm, contradict, and complement the statements made in previous studies mentioned in Section I and II.

Since implementing industrial symbiosis is highly context-dependent, small samples are not representative of a holistic understanding of the relevance of managerial and informational problems. Therefore, in order to reach as many actors as possible and to get a deeper and more holistic picture of these issues, we designed an online questionnaire targeting the relevant players of industrial symbiosis in industrial parks. By conducting the online questionnaire as the next step in our long-term research project, we aim to expand understanding of the problems/barriers that we have identified in the existing literature on industrial symbiosis, as discussed in Section II.

At this point in our research, the responses to the questionnaire are particularly important in providing guidance on how to design and use applicable ICT support in industrial parks and how to coordinate industrial symbiosis activities (e.g., information flow). After the results of the questionnaire have been obtained, the next steps in our research project will involve discussing the related issues in more detail with companies in industrial parks and with park management, using qualitative approaches such as interviews and focus group discussions. After that, we will use the results from both the questionnaire and the qualitative methods to develop an adequate, appropriate concept for an ICT support tool for industrial symbiosis for industrial parks, along with an instantiation of the tool. For the final step in our long-term research, the developed tool will be evaluated by companies in industrial parks and by park management, again using a questionnaire.

APPENDIX

**\*Please assess the current readiness of your company for industrial symbiosis in terms of business models and strategic orientation, regardless of whether cooperation opportunities actually exist.**  
 Your company (management) is ready to ...

	Strongly disagree 1	Disagree 2	Neutral 3	Agree 4	Strongly agree 5
replace raw materials with recycled materials and/or other waste and by-products from another company.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
provide by-products or waste to other companies as raw materials.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
invest in or share utilities or infrastructure with other companies.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
use or coordinate joint services such as disposal, procurement, logistics or training concepts.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
provide personnel and time resources for the activities of Industrial Symbiosis.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
align business strategy and corporate policy towards sustainability and industrial symbiosis.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
share the same goals with other companies or to develop and pursue common goals.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
be part of a network that is strategically oriented towards industrial symbiosis and sustainability.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Fig. 5 Screenshot: Example Survey Question – Readiness

**\*Which communication channels do you mainly use in your company to exchange experiences and knowledge with colleagues and partners?**

	Within your company					Outside your company				
	Never 1	Rarely 2	Some-times 3	Often 4	Always 5	Never 1	Rarely 2	Some-times 3	Often 4	Always 5
Face-to-face communication (e.g., workshops)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Virtual face-to-face communication (e.g., online)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
E-mail	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Intranet	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Expert systems / knowledge-based systems	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Fig. 6 Screenshot: Example Survey Question – Knowledge Sharing

**\*Are the following types of information classified as confidential in your company and therefore subject to disclosure restrictions to other companies (e.g., non-disclosure agreements)?**

	Raw materials			Waste and by-products		
	Yes	Uncertain	No	Yes	Uncertain	No
Resource type	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Resource quantity	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Supply pattern	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Resource properties	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Resource source	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Availability period	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Supplier/customer	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Price/cost	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Fig. 7 Screenshot: Example Survey Question – Information Sharing

## REFERENCES

- [1] M. R. Chertow, "Industrial symbiosis: Literature and Taxonomy," *Annual Review of Energy and the Environment*, vol. 25, no. 1, pp. 313–337, Nov. 2000.
- [2] D. R. Lombardi and P. Laybourn, "Redefining Industrial Symbiosis," *JIE*, vol. 16, no. 1, pp. 28–37, Feb. 2012.
- [3] E. A. Lowe and L. K. Evans, "Industrial ecology and industrial ecosystems," *Journal of Cleaner Production*, vol. 3, no. 1–2, pp. 47–53, Jan. 1995.
- [4] H.-S. Park, E. R. Rene, S.-M. Choi, and A. S. F. Chiu, "Strategies for sustainable development of industrial park in Ulsan, South Korea—From spontaneous evolution to systematic expansion of industrial symbiosis," *Journal of Environmental Management*, vol. 87, no. 1, pp. 1–13, Apr. 2008.
- [5] D. Sakr, L. Baas, S. El-Haggar, and D. Huisinigh, "Critical success and limiting factors for eco-industrial parks: global trends and Egyptian context," *Journal of Cleaner Production*, vol. 19, no. 11, pp. 1158–1169, Jul. 2011.
- [6] A. Golev, G. D. Corder, and D. P. Giurco, "Barriers to Industrial Symbiosis: Insights from the Use of a Maturity Grid," *Journal of Industrial Ecology*, vol. 19, no. 1, pp. 141–153, Feb. 2015.
- [7] R. Lombardi, "Non-technical barriers to (and drivers for) the circular economy through industrial symbiosis: A practical input," *Economics And Policy Of Energy And The Environment*, Feb. 2017.
- [8] G. B. Grant, T. P. Seager, G. Massard, and L. Nies, "Information and Communication Technology for Industrial Symbiosis," *JIE*, vol. 14, no. 5, pp. 740–753, Oct. 2010.
- [9] G. van Capelleveen, C. Amrit, and D. M. Yazan, "A Literature Survey of Information Systems Facilitating the Identification of Industrial Symbiosis," in *From Science to Society*, P. Hitzelberger, S. Naumann, V. Wohlgemuth, and B. Otjacques, Eds. Springer, Cham, 2018, pp. 155–169.
- [10] A. Maqbool, F. Mendez Alva, and G. Van Eetvelde, "An Assessment of European Information Technology Tools to Support Industrial Symbiosis," *Sustainability*, vol. 11, no. 1, p. 131, Dec. 2018.
- [11] D. Gibbs and P. Deutz, "Reflections on implementing industrial ecology through eco-industrial park development," *Journal of Cleaner Production*, vol. 15, no. 17, pp. 1683–1695, Nov. 2007.
- [12] E. Susur, A. Hidalgo, and D. Chiaroni, "A strategic niche management perspective on transitions to eco-industrial park development: A systematic review of case studies," *Resources, Conservation and Recycling*, vol. 140, pp. 338–359, Jan. 2019.
- [13] M. Chertow and J. Park, "Scholarship and Practice in Industrial Symbiosis: 1989–2014," in *Taking Stock of Industrial Ecology*, Springer, Cham, 2016, pp. 87–116.
- [14] S. Menato, S. Carimati, E. Montini, P. Innocenti, L. Canetta, and M. Sorlini, "Challenges for the adoption of industrial symbiosis approaches within industrial agglomerations," in *2017 International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, Funchal, 2017, pp. 1293–1299.
- [15] I. Siskos and L. N. V. Wassenhove, "Synergy Management Services Companies: A New Business Model for Industrial Park Operators," *Journal of Industrial Ecology*, vol. 21, no. 4, pp. 802–814, Aug. 2017.
- [16] L. R. Bacudio et al., "Analyzing barriers to implementing industrial symbiosis networks using DEMATEL," *Sustainable Production and Consumption*, vol. 7, pp. 57–65, Jul. 2016.
- [17] M. A. B. Promentilla et al., "Problematic approach to analyse barriers in implementing industrial ecology in philippine industrial parks," *Chemical Engineering Transactions*, vol. 52, pp. 811–816, 2016.
- [18] D. Ceglia, M. C. S. de Abreu, and J. C. L. Da Silva Filho, "Critical elements for eco-retrofitting a conventional industrial park: Social barriers to be overcome," *Journal of Environmental Management*, vol. 187, pp. 375–383, Feb. 2017.
- [19] U. Fric and B. Rončević, "E-Simbioza – Leading the Way to a Circular Economy through Industrial Symbiosis in Slovenia," *SocEkol*, vol. 27, no. 2, pp. 119–140, 2018.
- [20] M. Kurdve, C. Jönsson, and A.-S. Granzell, "Development of the urban and industrial symbiosis in western Mälardalen," in *Procedia CIRP*, 2018, vol. 73, pp. 96–101.
- [21] M. Ormazabal, V. Prieto-Sandoval, R. Puga-Leal, and C. Jaca, "Circular Economy in Spanish SMEs: Challenges and opportunities," *Journal of Cleaner Production*, vol. 185, pp. 157–167, Jun. 2018.
- [22] T. Domenech, R. Bleischwitz, A. Doranova, D. Panayotopoulos, and L. Roman, "Mapping Industrial Symbiosis Development in Europe: typologies of networks, characteristics, performance and contribution to the Circular Economy," *Resources, Conservation and Recycling*, vol. 141, pp. 76–98, Feb. 2019.
- [23] L. Kosmol, "Sharing is Caring - Information and Knowledge in Industrial Symbiosis: A Systematic Review," presented at the Conference on Business Informatics, Moscow, Russia, 2019.
- [24] M. R. Chertow, "'Uncovering' industrial symbiosis," *JIE*, vol. 11, no. 1, pp. 11–30, 2007.
- [25] G. Massard and S. Erkman, "A regional Industrial Symbiosis methodology and its implementation in Geneva, Switzerland," 2007.
- [26] L. Fraccascia, M. Mango, and V. Albino, "Business models for industrial symbiosis: a guide for firms," in *Procedia Environmental Science, Engineering and Management*, Italy, 2016, vol. 3, pp. 83–93.
- [27] F. Boons, W. Spekkink, and Y. Mouzakitis, "The dynamics of industrial symbiosis: a proposal for a conceptual framework based upon a comprehensive literature review," *JCP*, vol. 19, no. 9–10, pp. 905–911, 2011.
- [28] M. R. Ghali, J.-M. Frayret, and J.-M. Robert, "Green social networking: Concept and potential applications to initiate industrial synergies," *Journal of Cleaner Production*, vol. 115, pp. 23–35, 2016.
- [29] L. Fraccascia and D. M. Yazan, "The role of online information-sharing platforms on the performance of industrial symbiosis networks," *Resources, Conservation and Recycling*, vol. 136, pp. 473–485, 2018.
- [30] L. Kosmol and W. Esswein, "Capturing the Complexity of Industrial Symbiosis," in *Advances and New Trends in Environmental Informatics*, H.-J. Bungartz, D. Kranzlmüller, V. Weinberg, J. Weismüller, and V. Wohlgemuth, Eds. Springer International Publishing, 2018, pp. 183–197.
- [31] M. Benedict, L. Kosmol, and W. Esswein, "Designing Industrial Symbiosis Platforms – from Platform Ecosystems to Industrial Ecosystems," in *PACIS 2018*, Yokohama, Japan, 2018, pp. 26–30.
- [32] F. A. Halstenberg, K. Lindow, and R. Stark, "Utilization of Product Lifecycle Data from PLM Systems in Platforms for Industrial Symbiosis," *Procedia Manufacturing*, vol. 8, pp. 369–376, Jan. 2017.
- [33] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A Design Science Research Methodology for Information Systems Research," *Journal of Management Information Systems*, vol. 24, no. 3, pp. 45–77, Dec. 2007.
- [34] D. C. A. Pigosso, A. Schmiegelow, and M. M. Andersen, "Measuring the Readiness of SMEs for Eco-Innovation and Industrial Symbiosis: Development of a Screening Tool," *Sustainability*, vol. 10, no. 8, p. 2861, Aug. 2018.
- [35] F. Ceceja et al., "e-Symbiosis: technology-enabled support for Industrial Symbiosis targeting Small and Medium Enterprises and innovation," *Journal of Cleaner Production*, vol. 98, pp. 336–352, Jul. 2015.

# On the Use of Predictive Models for Improving the Quality of Industrial Maintenance: an Analytical Literature Review of Maintenance Strategies

Oana Merkt

Hohenheim University

Chair of Business Informatics II (530D)

Schwerzstrasse 35

Stuttgart 70599, Germany

Email: Oana.Merkt@uni-hohenheim.de

**Abstract**—Due to advances in machine learning techniques and sensor technology, the data driven perspective is nowadays the preferred approach for improving the quality of maintenance for machines and processes in industrial environments. Our study reviews existing maintenance works by highlighting the main challenges and benefits and consequently, it shares recommendations and good practices for the appropriate usage of data analysis tools and techniques. Moreover, we argue that in any industrial setup, the quality of maintenance improves when the applied data driven techniques and technologies: (i) have economical justifications; and (ii) take into consideration the conformity with the industry standards. In order to classify the existing maintenance strategies, we explore the entire data driven model development life cycle: data acquisition and analysis, model development and model evaluation. Based on the surveyed literature we introduce taxonomies that cover relevant predictive models and their corresponding data driven maintenance techniques.

## I. INTRODUCTION

THE quality of maintenance is a relevant aspect in the assessment of any industrial product or process, and therefore a challenging research problem. Our survey shows that maintenance approaches are continuously evolving over time. Earlier, corrective maintenance also known as reactive maintenance was used. Preventive maintenance proves to be a better alternative, as the maintenance actions are employed before the failure occurs. This approach evolved into condition-based maintenance, where decisions are based on the evaluation of the machine status through inspections and measurements. Among all the existent approaches to maintenance, each of them varying in terms of efficiency and complexity, predictive maintenance seems to best fit the needs of a highly competitive industry setup, as argued by [1]. Predictive maintenance allows maintenance actions to be based on changes of the machine and process parameters, which are continuously monitored by sensors. Currently, due to recent advances in sensor technology, data communication, and computing, the ability to collect big volumes of

heterogeneous, raw sensor data produced by equipment under observation is exponentially increasing. Therefore, historical information about normal and abnormal patterns and the related corrective actions employed during the lifetime of an industrial asset is becoming available. In order to deal with such high-dimensional problems, the predictive maintenance strategy uses a variety of techniques and prediction models that study both live and historical information. Further on, this information is used to learn prognostics data and to make accurate diagnostics and predictions, as presented by [2], [3], and [4]. They argue that the implementation of effective prognosis for maintenance has a variety of benefits including increased system safety, improved operational reliability, reduced maintenance, inspection times, repair failures and life cycle costs. Past works on predictive maintenance show that maintenance actions are performed by employing various prediction models and modeling techniques. Among prediction models, the machine learning (ML) approaches are typically considered the most suitable to deal with high dimensional and unstructured data, as argued by [8] and [9]. Moreover, multimodal fusion techniques are increasingly used by ML models for combining data from multiple, diverse modalities and sources with the goal of retrieving new insights from the fused knowledge i.e. multiple sensors may collect complementary or concurrent information which is fused in order to obtain more accurate machine diagnosis and prognosis. There is a lot of previous work on data fusion, as the topic dates back in the 90es. Application scenarios that implement ML models and apply multimodal data fusion for maintenance optimization purposes are defined by [2], [3], [9] and [10]. However, to date, no standard, nor good practice recommendations for fusion and integration of multimodal data have emerged. Our research work reviews the model-agnostic data fusion techniques in order to find solutions for their optimal usage. We argue that understanding the capabilities and challenges of existing multimodal data fusion methods and techniques has the potential to deliver better data analysis tools across all domains, including in the maintenance quality and management field of research.

This work was partly supported by a grant from the German Federal Ministry for Economic Affairs and Energy (BMWi) for the Platona-M project under the grant number 01MT19005D.

### A. Maintenance Issues Relative to Prediction Quality

We envision the problematic of maintenance quality as a complex topic with many complementary aspects: economical, the conformity with the mainstream industrial standards and technical. The first aspect follows the classical optimization concerns relative to maintenance costs, by considering aspects related to maintenance investment costs and resulting benefits.

Traditional approaches consider maintenance only as cost related. However, the maintenance activities have direct implications to the production and quality, therefore should be treated as an investment, as argued by [11]. Moreover, appropriate timing for performing maintenance activities has economical justifications, as explained by [12] in the description of the damage model. The damage model recommends the usage of maintenance actions only when clear evidence about the machine or equipment status exists. It shows that based on the long-term, historical data, it is possible to adapt the predictive maintenance interval to the industrial item life cycle by forecasting the items wear, and the impact of it on the production chain, respectively. Reference [12] explains that the probability of an item to fail is high at the beginning of its operational life, in its burn-in period. During the burn-in period, the failure probability of an item is constantly decreasing. During the items working period, the failure probability is low and remains constant, therefore the prediction of the items failure during the working period is challenging. The probability of failure raises with the working hours so that in the wear period the probability for an item to fail is again high. Therefore, [12] recommends as a good practice to perform maintenance actions during the wear period of an items life cycle.

The second aspect which, we believe, influences the quality of maintenance is conformity with industrial standards during the development life cycle of a maintenance product. Our review of the literature shows the problematic of ad-hoc maintenance model development and implementations that do not comply with the existing mainstream standards. This situation leads to the absence of good practice recommendations or general solutions in the development of maintenance products. We briefly review two existing industrial standards for model development: Cross Industry Standard Process for Data Mining (CRISP-DM) and Industry Data Space [13]. CRISP-DM standard represents a guideline to follow in the process of prototyping a learning model for maintenance purposes. We shortly list the guideline steps i.e. business understanding, data understanding, data preparation, data fusion, model prototyping, model evaluation, and deployment. A complete description is provided by the reference [45]. On its turn, Industry Data Space standard represents the solution to the actual problems raised by the huge volume of heterogeneous data which need to be handled in a standardized way in the industrial setup, as defined by [13]. Among the expected benefits of any standard, we mention the knowledge sharing and re-use which helps building complex, operational models.

The technical aspect of maintenance quality is related to

the set of decisions concerning the appropriate techniques and approaches that should be used for the development of an operational and highly qualitative maintenance model. Our literature survey mainly focuses on analyzing the technical aspect, but it considers also its connections with the economic aspect. To our knowledge, none of the reviewed research works takes into account the conformity with industrial standards for model development and data management and security. One of the main issues of actual maintenance techniques and methods is exactly the absence of this holistic view in considering the problem of the maintenance quality as directly influenced by all the above three mentioned aspects. The rest of the paper is structured as follows: Section 2 starts with a review of maintenance approaches, according to the terminology defined by both [14] and [15] maintenance standards. We introduce a taxonomy that covers the surveyed approaches by categorizing the employed predictive models, the corresponding modeling techniques and the implementation algorithms, respectively. Further on, we review the literature works focusing on the technical steps of a maintenance model development process: data acquisition and analysis, data fusion, model development and evaluation, each of them being discussed in a subsection. Moreover, we present the concept of multimodal data fusion and we discuss a taxonomy of model-agnostic data fusion methods and their usage recommendations. Section 3 presents the review process we followed in gathering the literature for our survey. Findings and results of the investigated approaches are highlighted in Section 4. Finally, Section 5 concludes the paper with a discussion about the research challenges and future works.

## II. BACKGROUND

### A. Classification of Maintenance Approaches

The European recognized maintenance standards: DIN EN 13306 - Maintenance Terminology [14], and DIN EN 31051 - Fundamentals of Maintenance [15] are defining the maintenance related terminology and concepts. According to the DIN EN 31051 standard, the maintenance concept is defined as: *the combinations of all technical and administrative actions as well as actions of management in the lifetime of a unit, in order to be in the fully functional state or to recover in this one, so that this unit can fulfill his requirements.*

The main maintenance activities i.e. service, inspection, repair, and improvement are defined by the DIN EN 31051 standard. Their definitions together with other relevant maintenance concepts defined by the DIN EN 31051 maintenance standard are listed in Table I. The DIN EN 13306 maintenance standard defines the existing maintenance strategies: *corrective maintenance, preventive maintenance, condition-based maintenance, and predictive maintenance.* They are discussed in the following subsections. Moreover, the definition of a further maintenance strategy, namely *prescriptive maintenance* - which is not yet standardized, but already used in practice - is discussed in the following subsection.

TABLE I  
FUNDAMENTALS OF MAINTENANCE DIN EN 31051 STANDARD

Item	Defines a component, device, subsystem, functional unit, equipment or a system which can be described and considered as an entity.
Wear	Represents the reduction of wear margin due to chemical or physical processes.
Wear limit	Is the defined minimum value of the wear margin.
Wear margin	Defines the possible reserve function capacity under defined circumstances which a unit possesses.
Service	Includes all activities delaying the degradation of the wear margin. The activities include cleaning, conservation, greasing, oiling, complementing, changing and readjusting.
Inspection	Refers to all activities used to determine and evaluate the actual conditions of facilities, machines, assemblies or components. Inspection refers to collecting data, and related activities that can be measuring, verifying and monitoring.
Repair	Covers activities for retrieving the nominal condition, such as renewing, patching and adjusting.
Improvement	Defines the combination of all technical and administrative activities as well as activities of management in order to increase the reliability, the maintainability, or the safety of an item without changing its initial function.

1) *Corrective Maintenance*: According to the EN 13306 standard, the corrective maintenance is defined as *the maintenance carried out after fault recognition and intended to put an item into a state in which it can perform a required function*. A system that employs corrective maintenance is aware of all its predefined set of failures and damages. But, in the industrial, operational context new faults and their corresponding patterns appear over time, because of the items usage during the working hours. One main advantage of applying corrective maintenance techniques is that the wear-limit of an item, i.e. the service time is fully used. This implies that the effort for items inspection and for replacing the item is significantly reduced, compared with the case of preventive maintenance. The main challenge in applying corrective maintenance is that the item can fail at an unknown time not previously known or decided and consequently can produce damages and an additional cost that can be higher as the yield of full usage of its wear margin.

2) *Preventive Maintenance*: The EN 13306 standard defines preventive maintenance as *the maintenance carried out at predetermined intervals or according to prescribed criteria and intended to reduce the probability of failure or the degradation of the functioning of an item*. One main challenge of preventive maintenance in operational context is that industrial scenarios for data analysis do not provide tracking of the past, abnormal behavior or maintenance operations that were performed in order to correct or to prevent a faulty behavior. Consequently, preventive maintenance defines a set of actions carried out before failure and that are intended to prevent failures or degradation of a machine. Time-based maintenance is defined as the preventive maintenance approach that recommends performing all maintenance activities after a certain amount of operation hours, or by predefined scheduling, regardless of the items health condition. The assumption is that after several operational hours, the wear margin of an item is worn out. The employed approach is to change the item or to overhaul part of it before the wear margin is used. The advantages of time based maintenance are the reduced breakdown frequency and the increased service life compared with other preventive maintenance strategies. It is therefore recommended only when the safety of the environment can be harmed, or when the items lifetime is known, which is not the case in the operational environment. The economic justi-

fication behind the time-based maintenance approach is that the maintenance costs can be kept low when the maintenance interval is adjusted to the actual lifetime of the asset so that the item or some of its parts are changed just before they fail.

3) *Condition-based Maintenance*: The EN 13306 standard defines condition-based maintenance as *the preventive maintenance which includes a combination of condition monitoring and/or inspection and/or testing, analysis, and the ensuing maintenance actions*. Condition-based maintenance aims to anticipate a maintenance operation based on the evidence of degradation and deviations from a supposed asset normal behavior. The equipment is monitored with multiple sensors which are supposed to acquire relevant data about the equipment operation life. Additionally, contextual information like temperature, humidity, etc. may also provide significant information. Key Process Indicators (KPIs) or health indicators are usually computed and analyzed, in order to discover trends that lead to abnormal contexts and failure events.

4) *Predictive Maintenance*: According to the EN 13306 standard, predictive maintenance is defined as *the condition-based maintenance carried out following a forecast derived from repeated analysis or known characteristics and evaluation of the significant parameters of the degradation of the item*. Predictive maintenance is a sub-class of condition based maintenance. It uses a variety of approaches and ML techniques to study both recent and historical data and to learn prognostic models which are expected to make accurate predictions about the future status of a machine or equipment. The main challenge of predictive models is that they rely on the assumption that there are certain contexts in the equipment life time where the failure rate is increasing. In the industrial, operational context there are patterns in which the failure probability does not increase, but remains constant during the equipment life time, and therefore the equipment can fail at any time: it is the case of electrical and electronic components.

5) *Prescriptive Maintenance*: Terminologically, it is mentioned by neither the EN 13306, nor the DIN EN 31051 maintenance standards. But, its functionality can be consequently deduced and is seen as a recommendation of one or more courses of action based on the outcomes of models for corrective and predictive maintenance. The main challenge of prescriptive maintenance is the difficulty to build in practice operative models. Existing research models are based on ad-

hoc model development where ML methods and data fusion techniques are jointly used with fuzzy reasoning, simulation techniques, and evolutionary algorithms.

Tables II, III and IV introduced in Section 4 are constructed based on the reviewed literature on maintenance strategies i.e. corrective, preventive and predictive. The tables present the surveyed literature, i.e. a structured review of the maintenance type and goals, correlated with a specific statistical or data-driven operational method, and the corresponding results. For a better understanding of implementation techniques for maintenance purposes, the next section reviews the basic steps of a data-driven model development life cycle i.e. data acquisition and preparation, model development (including the multimodal ML methods discussion) and model evaluation.

### B. Data Driven Model Development Life Cycle Methods and Techniques for Maintenance Purposes

Understanding the specific application context, or the business requirements is the first step for any learning model developed and deployed in an industrial environment. The basic steps of a data driven model development life cycle for maintenance purposes are discussed in the next subsections.

1) *Data Acquisition and Preparation*: Predictive models learn patterns from historical, multimodal data and predict future outcomes with certain probability based on these observed patterns. The performance of any learning model is highly correlated with the relevancy, sufficiency, and quality of the training, validation and test data. *Data pre-processing* and *feature extraction* techniques are relevant in building reliable data driven models. Processing the raw data before modeling is improving the performance of the learned model. In practice, raw data in the form of sensor signals are complex and related information about the degradation process of the monitored component is not always available. Therefore, *pre-processing* raw sensor data is a mandatory step before building the maintenance models. Generally, data processing methods can be divided into two main tasks, namely *processing* and *data analysis*.

2) *Model Development*: In the context of a data-driven model development life cycle, the ML techniques for maintenance is considered the most suitable research perspective to deal with big volumes of heterogeneous data. ML techniques comprise two main approaches: (i) *supervised learning*, where the information about the occurrence of failures is present in the modeling data set; and (ii) *unsupervised learning*, where only the process information is available and no historic maintenance data exists. In an operational environment, predictive maintenance makes use of the following well-established supervised learning techniques from ML field: (i) *classification algorithms* which are used to represent groups of normal and abnormal health status of the item under observation i.e. Random Forest, Nearest Neighbors, SVMs and HMMs; (ii) *regression algorithms*; and (iii) *clustering methods* with anomaly detection algorithms. Multimodal machine learning (MML) represent an increasingly used set of ML methods for combining data from multiple and diverse modalities and

sources with the goal of retrieving new insights from the combined knowledge i.e. multiple sensors collect complementary or concurrent information which is combined in order to obtain more accurate machine diagnosis and prognosis. We provide in Section 4 an overview of the multimodal ML methods which shows that the multimodal fusion method seems to be the most employed for maintenance goals.

3) *Model Evaluation*: Once a model is built, an estimate of its performance is required. According to [41] there are two types of evaluation metrics that are giving insights about the quality of the model performance metrics: offline evaluation metrics that measure offline data of the prototype model and uses mainly historic data, and online evaluation metrics that measure live metrics on the deployed model on real-time data. Offline evaluation is used to estimate the performance of training and validating data, and therefore performance metrics like accuracy and precision-recall together with F1-Score are employed. An online evaluation usually is used for real-time, test data e.g. to estimate business metrics. Our survey focuses on reviewing the offline evaluation metrics used in industry for evaluation learning models for prediction maintenance purposes. The model evaluation is made on a different set of data, i.e. testing data set that is statistically independent of the data set that it was previously trained on. Mathematically speaking, the model evaluation means to estimate the generalization error of the learning model, i.e. to measure how good the model behaves under new data calibrations. A good practice is to split the data set into training, validation and test data, in a time dependent manner. Further good practice is to consider the training data earlier in time than all the validation and test data. Andrew Ng recommends a split such as training set (60%), cross-validation-set (20%) and testing set (20%). The confusion table (also called confusion matrix) it is used to show a detailed breakdown of correct and incorrect classifications and is applied for evaluating models that learn from highly imbalanced data. Performance metrics based on the confusion table are accuracy, precision, recall, specification, F1-Score, the AUC-ROC Curve.

## III. RESEARCH APPROACH

The aim of this research review is to increase knowledge in the field of maintenance techniques and their operationalization. This implies a sort of awareness in considering the appropriate predictive models and their corresponding implementation techniques depending on the available data and on the application scenario. The conference and journal publications selected for our review belong to the non-empirical conceptual and mathematical field of research. Consequently, they describe issues and perspectives related to maintenance strategies and their modeling techniques applied in an industrial setup. A systematic search using online databases was employed for a keyword-based search, in order to find journal and proceeding publications. We used the English language and the following keywords: *maintenance AND machine learning*. We iteratively continued the search using the following keywords: *predictive*

*maintenance, multimodal machine learning, multimodal fusion, multimodality, maintenance AND big data, maintenance AND Industry 4.0.* We finally obtained a shorter literature list which was further reduced by eliminating the duplicates, when similar topics and approaches were found. Science Direct, Scopus and Google Scholar literature databases were used for their wide coverage of journals, proceedings and books.

#### A. Description of the Criteria Used for Analysis

Our research perspective relative to the maintenance quality problematic focuses on: (i) the decision process to choose a specific maintenance approach i.e. maintenance goals, benefits, challenges and obtained results; and (ii) the implementation of the maintenance approach i.e. the employed prediction models and their corresponding modelling techniques. The selected literature was carefully examined in order to extract useful information based on the following criteria:

- *Prediction models* reveal a taxonomy of the most employed prediction models types employed in a maintenance process i.e. physical models, knowledge-based models, databased models and hybrid models.
- *Modelling techniques* represent the implementation pipeline (data analysis + algorithms) used. It is a relevant criterion which further helps us to select the set of the most used ML algorithms to be critical reviewed.
- *Dataset* comprises information about the involved sensor types and the kind of fusion applied. It is a relevant criterion which further helps us to provide a critical analysis of the quality of data involved in a maintenance process.
- *Industry* is concerned with the branch of industry where maintenance processes are applied.
- *Equipment parts* reveal the critical parts of equipment which are considered for maintenance.
- *Obtained results /performance metrics* extract the information concerning how the model was evaluated and give us a hint about how optimal the data analysis and learning algorithms were applied.
- *Maintenance goals* provide us with a taxonomy of topics showing the final decisions of the algorithms pipeline. Paired with the Modelling techniques criterion, it gives useful information about the successful algorithm pipeline used for a certain maintenance goal.

The overview of the reviewed maintenance literature is presented in Tables II, III and IV. We are not considering for our research works the empirical perspective, i.e. we are not discussing the maintenance strategies and their operationalization based on information obtained from interviews, or from analyzing case studies. The analytic literature review we conduct is formalized by [17] and [18] and starts with clarifying relevant maintenance terminology and definitions based on the accepted, European maintenance standards [15] and [14].

## IV. FINDINGS AND RESULTS

This section presents the reviewed results displayed in Tables II, III and IV. The surveyed works we consider are grouped by maintenance type, and further on they are grouped by prediction modeling types and relevant modeling techniques used in the implementations.

### A. Analysis of Maintenance Strategies

1) *Corrective Maintenance:* Our survey shows that the *fault recognition and diagnostic* is generally seen as a process of pattern recognition i.e. the process of mapping the information i.e. the features obtained in the measurement space to the machine faults in the fault space, as described in [19], [20], [21] and [22]. Diagnosis is a necessary part of any maintenance system, as only prognostics cannot provide in practice a sure prediction which covers all failures and faults. In case of unsuccessful prognosis, a diagnosis is a complementary tool for providing maintenance decision support. The methods employed in order to deal with fault classification and diagnostics are diverse: from expert systems [23] to Hidden Markov Models (HMM)s, as presented in [19], Artificial Neural Networks (ANN)s as described in [20], Support Vector Machine (SVM) as in [21] and fuzzy algorithms enhanced with spectral clustering and Haar wavelet transform, as described in [22].

2) *Preventive Maintenance:* The reviewed literature shows that a relevant class of preventive maintenance techniques are the prognostics through pattern recognition, classification and machine health status identification. Prognostics analyze data by automatically finding new insights in terms of behavioral patterns. The information extracted from the monitored data can help detecting patterns that characterize the machine working conditions or is anticipating and estimating critical events i.e. fault detection as in [3] and Remaining Useful Life (RUL) estimation as in [8]. Prognostics are considerate superior to diagnostics in the sense that they prevent faults and are employed for prediction problems with items spare parts and human resources, saving unplanned maintenance costs. The reference [5] proposes a data mining maintenance approach for predicting material requirements in the automotive industry by measuring the similarity of customer order groups. Identifying behavioral patterns in data means to classify similar data in some data-groups which share the same characteristics i.e. operational conditions, as described by [24], [25], [26], [27] and [28]. Within these classified groups there are data-points that are far from the identified pattern i.e. the outliers, or they may correspond to a distinctive property i.e. the mean point or the group distribution. Such patterns may help to identify faults or any other type of abnormal behavior. Large groups of data are interpreted as normal behavior, while small groups of data, or events that are far from the pattern are usually representing anomalies. Consequently, in modeling the learning model, there are only unlabeled input examples, i.e. we employ the the unsupervised learning perspective. ML algorithms and data fusion strategies are used to find new patterns in data therefore, in this case, clustering should be the most used technique,

TABLE II  
REVIEW OF CORRECTIVE MAINTENANCE MODELS AND CORRESPONDING IMPLEMENTATION TECHNIQUES

References	Prediction models	Modeling techniques	Dataset(s)	Industry	Equipment parts	Obtained results e.g. performance metrics	Maintenance goals
(Bunks, C., et al., 2004) [19]	Knowledge based models	Expert Systems + fault tree analysis	functional sensor data of the PV pilot plant + meteorological sensor data	Water pumping station	PV pilot plant Zambelli, Italy (Joule II EU Project)	real time supervision and monitoring + detection of foreseen faults	real time monitoring; maintenance inspection on request
(Deuszkiewicz, P., et al.; 2003) [20]		Fuzzy similarity, fuzzy c-means algorithm	synthetic data of simulated faults in a pressurizer water reactor (PWR) NPP: vapor and steam temperature, liquid temperature, liquid level and pressure	Nuclear Power Plant	Pressurizer water system	drawback: new faults cannot be classified into new groups without repeatedly applying the spectral analysis	classification models for fault diagnosis using unsupervised clustering
(Hao, Y., et al.; 2005) [21]	Data based models	Stochastic model: Hidden Markov Models (HMMs)	vibration measurements from a set of 8 accelerometers from the gearbox, at 9 torque levels and 8 seeded defects	Naval Research	Westland helicopter gearbox	HMMs are fully probabilistic models incorporating quasi-stationarity as a feature + build robust and flexible classification models	machine health status diagnostics; defect type classification
(Baraldi, P., et al.; 2014) [22]		Artificial Neural Networks (ANNs)	sensors from the body of the driving axle box at various speed (50, 70, 90, 110 Km/h), crest factor signal, XSK signal	Railway (ZNTK S.A. Rolling Stock Repair Company)	Power transmission unit in (ED-72 train) rolling bearing	minimizes the frequency of revision inspections + in time online warning for unexpected new failures	Machine health status diagnostics in useful time
(Alexandru, A., 1998) [23]		Statistical model: Support Vector Machine (SVM)+ k-fold cross validation	gas temperature, fuel flow, pressure rotor speed	Aerospace	Gas Turbine engine	accuracy: 93% even when the standard deviation of noise is 3 times larger than normal: a better generalization than ANNs	Identification of 3 most possible faults types

TABLE III  
REVIEW OF PREVENTIVE MAINTENANCE MODELS AND CORRESPONDING IMPLEMENTATION TECHNIQUES

References	Prediction models	Modeling techniques	Dataset(s)	Industry	Equipment parts	Obtained results e.g. performance metrics	Maintenance goals
(Manco, G., et al.; 2017) [3]	Hybrid models	Outlier detection	failures, events described by type, timestamp, subsystems, duration, severity, description	Railway	Train doors	High degree outliers are effective indicators of incipient failures.	fault detection
(Krishnakumari, A., et al.; 2017) [24]	Knowledge based models	Fuzzy Classifier + Decision Tree	Feature extraction + monitored data representing condition-based status of vibration signals	Manufacturing	Gears in rotary machines	Feature extraction and classification explained. The performance of the fuzzy inference has 95 %accuracy.	pattern recognition + fault detection and classification
(Jaramillo, V.H., et al.; 2017) [25]		Statistical model: Bayesian Inference	Multi sensor feature based fusion (acceleration, current, voltage, temperature)	Manufacturing	Electric motor with two gearboxes and a load	Feature based fusion + concepts of global/local fusion + feature extraction is good explained based on the example + transparent Bayesian inference method	machine health status assessment and condition monitoring
(Liu, C., et al.; 2016) [26]	Data based models	Statistical model: SVM +Fourier transform + discrete Wavelet decomposition	Multisensor feature-based fusion (dynamometer sensor, acceleration sensor, cutting force, vibration signal)	Manufacturing	Cutting tools and flank milling machines	Accuracy: 90% information feature-based fusion with multiple sensors provide complementary information to machining conditions	multiple machine condition monitoring and recognition
(Diez, A., et al.; 2016) [27]		k-NN based outlier remover + clustering approach of vibration events and joints + Fourier transform	Multi-sensor feature-based fusion (accelerometer sensor + location sensor)	Construction	Bridges (Sydney Harbour Bridge)	Real time health score (of the structure) learned from historical data and used to check new events based on cluster centroids and joints representatives.	damage detection of abnormal or damaged
(Li,C., et al.; 2016) [28]		ANN and Deep Learning	Automatic multisensor feature fusion from vibration, signal measurements	Manufacturing	Rotary machines	Deep Learning with statistical feature representation shows better performance metrics. Statistical features in the time, frequency and time-frequency domains have different representation capabilities for fault patterns.	fault diagnostic and fault patterns identification

together with a measure of similarity which should deal with showing the correspondence of data groups. When pattern classification is applied for describing training data, then we assume the availability of (i) historic data with abnormal behavior; and (ii) data concerning maintenance activities that were carried out. The learner looks for identifying the causes for confirmed, abnormal behavior and critical events, in order to predict them and to avoid them in the future. In this case, values for the target labels are available, and therefore the supervised learning strategy and the corresponding algorithms are to be employed. The target labels are representing features that are discrete or continuous, and they are always related to the diagnostic. Time-series analysis is employed to extract damage and fault-sensitive features from data. When a corrective action is made, the preceding data represent an abnormal behavior or abnormal data context. When events of interest or based on past maintenance actions are tracked, they are assumed to represent a normal data context. Another type of scenario is learning the normal behavior of the machine or parts from its equipment. This is a difficult process in the operational context, as it supposes that there are no outliers, nor operational faults, which is not the case of the industrial environment. Finding patterns in the monitored data requires a deep knowledge of the topic and of physics of the process so

that the issue can be theoretically understood. In the context of supervised learning approaches, feature engineering and mainly the interpretation of the assessment of the results represent always a challenge.

3) *Predictive Maintenance*: The survey shows that the predictive maintenance process has the goal of providing an accurate estimate of the RUL, but also it should assess the provided estimate, as argued in [31], [32], and [33]. Time-series analysis is used to anticipate anomalies and malfunctions in equipment and processes maintenance procedures. Traditional approaches are moving average over a time window, ARMA/ARMAX, Kalman Filter and cumulative sum, as described in [9]. Recursive Neuronal Networks (RNNs) show relevant characteristics for time series forecasting, as their loops allow information to persist, as presented in [8]. Multi sensor fusion ranges from multi signal combinations, as argued in [8] and [9], to more complex integration of conditional assessment, RUL estimation, and decision making, as presented in [2] and [10]. Operational predictive approaches are based on a schema that implies frequent, and sometimes unnecessary maintenance of the equipment and of the entire production process that leads to high maintenance time and costs. They use complex A.I. based algorithms, and data fusion strategies - in an ad-hoc manner, usually after trial and error approaches - which imply the usage

of consecutive fusion algorithms, as described by reference [26]. The uncertainty in prediction is always a challenge and to this time the fuzzy logic is used to represent uncertainties in prediction, as argued by [4]. As a particular case of condition-based maintenance, reference [29] shows that techniques for condition monitoring and diagnostics are gaining acceptance in the industry sectors, as they prove to be effective also in the predictive maintenance and quality control areas. The authors apply a feature based fusion technique implemented with the cascade correlation neuronal network to multiple sensor data collected from rotating imbalance vibration of a test rig. The results show that the multi-sensory data fusion outperforms the single sensor diagnostic. The reference [30] focuses on the capability of providing real-time maintenance by extracting knowledge from the monitored assets (with vibration sensors) on the production line. Using intelligent data driven monitoring algorithms (ADMM), data fusion strategies and the proposed three-levels layered (IoT, Fog with gateway nodes for sensors aggregation, Decision) system model, the authors argue on the efficiency of cloud oriented maintenance.

The uncertainty in prediction is always a challenge and to this time the fuzzy logic is used to represent uncertainties in prediction, as argued by [4]. The references [6] and [7] show that the problem of scheduling under constraint of completion time of all production jobs can also be solved using predictive maintenance algorithms. The efficiency of the algorithms for predicting machine failures is further evaluated using simulation tests. The results, i.e. the optimized jobs schedule shows a nearly 50% drop in the number of operations compared with the initial, nominal schedule. The classification of modeling techniques for predictive models is presented in Figure 1. Physical models use the laws of physics to describe the behavior of a failure [2]. *Knowledge-based models* assess similarities among observed situations and a set of previously defined failures. These models can be sub-divided in *expert system models* able to answer complex queries, as presented by [23], and *fuzzy models*, as in [4]. *Data-driven models* are based on the acquired data. This type of model can distinguish among *stochastic models*, *statistical models* and *artificial neural networks* (ANNs). *Hybrid models* use combinations of two or more modeling techniques as in [34], [35] and [44]. Stochastic models provide event-based information. Hidden Markov models and Kalman filters belong to this category too. Statistical models predict a future state by comparing the monitored results with a machine health state without faults. ML models, such as regressions, classifications, and clustering represent a category of data-based, statistical models relevant in the study of maintenance optimization. However, the ML models are focusing on increasing the accuracy of their predictions, while the classical statistical community is more concerned with the understanding of their models and of the model's parameters i.e. model calibration and inference.

4) *Prescriptive Maintenance*: The main challenge of prescriptive maintenance is the difficulty to build in practice operative models. Existing research models are based on ad-hoc model development where ML methods and data fusion

techniques are jointly used with fuzzy reasoning, simulation techniques, and evolutionary algorithms. The reviewed literature shows that prescriptive maintenance implementations show an ad-hoc grouping of methods including data fusion and ML techniques, combined with fuzzy reasoning algorithms, simulations [34] and multi-objective evolutionary algorithms for optimization [44]. When a predictive model raises an alarm before the fault occurs, the prescriptive model will work in the direction of reducing the probability that this alarm will rise in the future, by modifying the working parameters and variables of the asset or the process affected by the fault. When the fault is confirmed, the prescriptive models will work to minimize its impact of the work context and to re-routing assets to the non-faulty production lines.

The tables II, III and IV are constructed based on the reviewed literature on maintenance types: corrective, preventive and predictive. The tables present a structured view of the maintenance type and goals, correlated with a specific statistical or data-driven operational method, and the corresponding results.

## B. Analysis of Data Driven Development Life Cycle

1) *Data Acquisition and Preparation*: Predictive models learn patterns from historical, multimodal data and predict future outcomes with certain probability based on the observed patterns. The performance of any learning model is highly correlated with the relevancy, sufficiency, and quality of the training, validation and test data. Moreover, the data used for training and testing the model should be relevant for the application scenario, therefore the expertise and the guidance of a domain expert is important. The most relevant data sources for a predictive model application scenario are condition monitoring data referred to as hard data, and human generated data referred to as soft data. Condition monitoring data contains knowledge in the form of degradation patterns and other types of anomalies in data that leads to an item degradation. Time-varying features are expected to capture these abnormal patterns, and the models fed with these features are expected to learn to distinguish between normal and abnormal pattern behaviors of items and also to forecast the RUL for the monitored items. Condition monitoring data can be further decomposed into sensor data, asset data, operation data, offline inspection data, and historical data. On its counterpart, human generated data represents information about replaced components, repair activities performed on a certain item or on parts of it. Moreover, it consists also of software generated information e.g. event data information such as alarms and faults messages which are described in natural language, but it comprises also technical metadata for devices and processes i.e. model, manufactured date, the start of service, maintenance reports. The event-data collection implies always a manual process and includes qualitative information about the monitored item such as the description of the installation, breakdown, inspection, repair, overhaul, failure causes, etc., the severity of the failure and the description of what was done to fix the failure. In practice, the item under critical

TABLE IV  
REVIEW OF PREDICTIVE MAINTENANCE MODELS AND CORRESPONDING IMPLEMENTATION TECHNIQUES

References	Prediction models	Modeling techniques	Dataset(s)	Industry	Equipment parts	Obtained results e.g. performance metrics	Maintenance goals
(Xenakis, A., et al.; 2019) [30]	Knowledge based models	Rule-based fuzzy logic + condition-based fusion diagnosis	Multisensor decision level fusion (vibration signal + current signal)	Railway	Electric multiple units (EMU) trains ->pulling motor of EMU bogie	the accuracy of multiple classifier fusion (vibration/current features) is greater as the accuracy of single classifiers	general maintenance
(Liu, Z., et al.; 2018) [2]	Data based models	ADMM (altering direction method of multipliers) algorithm + Decision Fusion	1-second vibration signals snapshots with the sampling rate set at 20kHz	Industrial Automation	Production Line	minimize operational costs + efficient energy consumption	real time analyse and process of machine faults + health status monitoring
(Niu, G., et al.; 2017) [4]		RNN-based health indicator for RUL prediction	Multisensor fusion at feature level (vibration signals + time-frequency features)	Aerospace	Bearings	high RUL prediction accuracy of generator bearings	RUL prediction
(Guo, L., et al.; 2017) [8]			KIP from operational data (NASA Lithium-Ion Battery B0005->B0056 repository)+ Turbofan engine data C-MAPSS	Aerospace	Battery and turbofan engine	3-fold cross validation is successfully validating the approach. average MAPE is computed and generates low errors for both applications	RUL prediction
(Acorsi, R., et al.; 2016) [9]			Features extraction from single product and fleet levels	Electrical Power plants	Medim/High circuits breakers	Health Condition Profile with RUL and PoF (Probability of Failure) computed in a predetermined window of time.	RUL and PoF prediction
(Mosallam, A., et al.; 2016) [31]			Multisensor fusion (waterfall fusion technique) at feature level + decision level (accelerometer data + load cell data)	Manufacturing	Rolling bearings	Data from different sensors provide more information than data gathered from single ones.	condition-based monitoring and diagnosis
(Cristaldi, I., et al.; 2016) [33]	Hybrid models	k-means, association rules (GSP, Apriori), Neural Networks, Random forest, Decision Tree, kNN	Feature extraction from parameter logs (user settable machine quantities), message logs and energy data sampling sensors	Manufacturing	Automatic machines in the manufacturing line	accuracy (95% - Random Forest), but the precision is low (38%) which implies false alarms recall (74% - Neural Networks)	fault prediction
(Safizadeh, M., et al.; 2014) [10]		Simulation + multi-sensor fusion	Multi sensor hard/soft data fusion	Aerospace	Aerospace Industry Manufacturing	digital twin concept and many levels of fusion for hard/soft data	health status estimation and maintenance

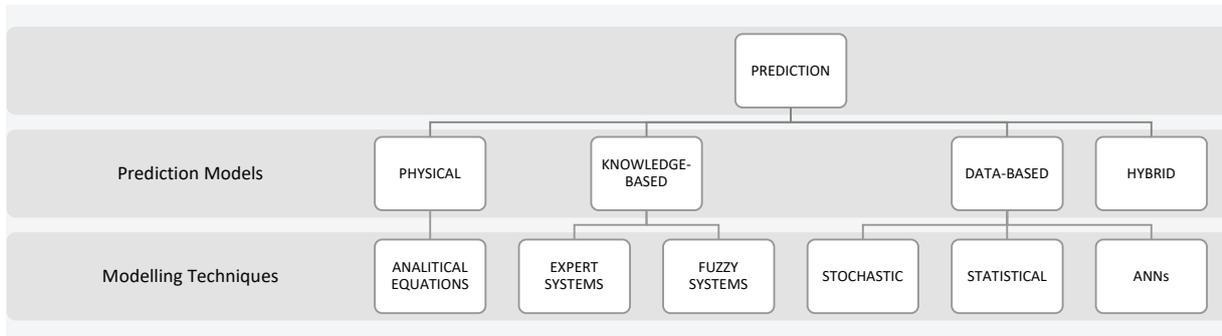


Fig. 1. Taxonomy of Prediction Models

conditions which is monitored continuously generates two types of data: event data and condition-monitoring data. Event data represent fault events which are considered critical to the system, and diagnostics messages when the events are alarm messages that described the item status. Events are triggered by the software component that monitors the item based on the item status information. Condition monitored data is collected every time when the events are triggered in order to form the context of the associated events and to ease their interpretation. Event data is characterized by attributes like type (i.e. fault, alarm, diagnostic), timestamp, item/sub-component where the event was triggered, severity, duration, and textual description, among many other possible attributes. After the acquisition process, the data sets must be pre-processed as they exhibit uncertainties that may affect the learning model performance. Data preprocessing and feature extraction techniques are relevant in building reliable data driven models. Processing the raw data before modeling is improving the

performance of the learned model. In practice, raw data in the form of sensor signals are complex and related information about the degradation process of the monitored component is not always available. Therefore, preprocessing raw sensor data is a mandatory step before building maintenance models. Generally, data processing methods can be divided into two main tasks, namely processing and data analysis. After processing the raw data coming from sensors, the resulting heterogeneous data may be categorized in the following types, depending on the quality of information they provide: (i) competitive, or redundant sensor data; (ii) cooperative, non-overlapping but partial sensor data; (iii) complementary, overlapping and partial sensor data; and (iv) independent, unrelated sensor data. Feature engineering is the next step prior to modeling the data. A feature is considered to be a predictive attribute for the model, such as temperature, pressure, vibration, etc. It is a good practice that the features extracted from the sensor data to comply with the following requirements: (i) features should

contain information required to distinguish between potential faults; (ii) features should not take into account the irrelevant variability which might be mixed in the sensor signals; and (iii) features should be limited in number to allow efficient computation.

2) *Model Development*: ML techniques for predictive maintenance comprise two main approaches: (i) supervised learning, where the information about the occurrence of failures is present in the modeling data set; and (ii) unsupervised learning, where only the process information is available and no historic maintenance data exists. In an operational environment, predictive maintenance makes use of the following well-established techniques from ML field: (i) classification algorithms which are used to represent groups of normal and abnormal health status of the item under observation: Decision Tree, Random Forest, Nearest Neighbors, SVMs and HMMs; (ii) regression algorithms; and (iii) clustering methods with anomaly detection algorithms, as presented in Figure 2. Binary classification algorithms are used to predict the probability that a piece of equipment fails within a future time period. The business requirements, the analyzed available data and the domain expert make estimation for e.g. (i) minimum lead time required to replace components, deploy resources and perform maintenance actions in order to avoid a problem that is likely to occur in the future time period; or (ii) minimum count of events that can be triggered before a critical problem occurs. Multi-class classification algorithms are used for making predictions in the following possible scenarios: (i) defining a plan maintenance schedule i.e. estimation of the time intervals when an asset has the bigger probability to fail; (ii) monitoring the health status of an asset i.e. estimation of the probability that an asset will fail due to a specific cause /root problem; and (iii) prediction that an asset will fail due to a specific type of failure. In this case, a set of prescriptive maintenance actions can be considered for each of the previously identified set of failures. Another type of algorithms for classification are the multiple classifiers which can be used in the process of knowledge discovery to discern particular patterns of data degradation for an asset or for a process. The benefits of the multiple classifiers reside in allowing the planning of the maintenance schedules using a statistical cost minimization approach. Regression models are typically used to compute the RUL of an item, as presented in [8]. RUL is defined as the amount of time that an asset is operational before the next failure occurs. The operational historical data is needed because the RUL calculation is not possible without knowing how long the asset has survived before a failure. Autoregressive models such as ARMA models assume that all future values are linear functions of past observations. e.g. fault predictions. A data-based ANN approach is recommended to be used for information clustering when there is no knowledge or understanding about the monitored system e.g. [8] and [10].

3) *Multimodal Machine Learning Methods*: Multimodality is defined by [36] as referring to the way something happens, or is experienced: we read textual information, we see objects and we hear sounds, we feel textures and smell odors. All

these perceptions represent modalities. A research problem, application or data set is multimodal when it includes multiple such modalities. In order to understand and to make sense of the world around us, A.I. techniques, in particular, multimodal machine learning (MML), must be able to interpret multimodal information and further to reason about it and make decisions. MML is a *multi-disciplinary field of research which builds models that process and relate information from multiple modalities*, as defined in [36]. The main idea is that *data from different sensor sources provide different representations of the same phenomena*. In MML literature, this is known as *multimodal, multi-view, multi-representation or multi-source learning*, as described in [37]. The main multimodal ML methods were identified and defined in [36] i.e. representation, translation, alignment, fusion, and co-learning. Their definitions according with [36] and [37] are listed in Table V. Understanding the capabilities and challenges of existing multimodal data fusion methods and techniques has the potential to deliver better data analysis tools across all domains, including the maintenance quality and management field of research. A relevant research challenge for the multimodal data fusion perspective is to identify patterns and commons governance rules that can be used to apply the appropriate multimodal data fusion technique for an application specific context or for a data set. Reference [38] arguments that data fusion is a multi-disciplinary research area with ideas raised from many diverse research fields such as signal processing, information theory, statistical estimation and inference, and artificial intelligence. Data fusion appeared in the literature as mathematical models for data manipulation. The diversity of the research fields is indeed reflected in the reviews of maintenance techniques in Tables II, III and IV. Multimodal data fusion represents the integration of information from multiple modalities, with the goal of (i) making a prediction; and (ii) retrieving new insights from the joined knowledge, as defined by [36]. There are many approaches to data fusion, as the topic dates back in the 90es. The model-agnostic technique to data fusion is discussed in [36] and [39] and later, described by [37], which also lays the grounds for the multimodal data fusion formal theory. Multimodal data fusion has the direct economic impact in the implementation of maintenance techniques which are based on aggregation data from heterogeneous sources into actionable decisions for maintenance purposes. Multimodal data fusion represents the core concept in MML, as argued in [36] [39]. The model-agnostic data fusion types that are used in the operational environment are listed in Table VI. The reference [37] lays the grounds for the multimodal data fusion theory by giving a solution to the research problem of determining the appropriate type of data fusion for a specific application context or for a data set. On his view, the main challenge in multimodal data fusion research resolves around the dependency-problem i.e. the arguments for choosing a specific type of data fusion. The assumption is that the optimal fusion type to be employed in an operational environment depends on the level we expect to see a dependency between the inputs in the modalities: (i) feature-based fusion assumes a

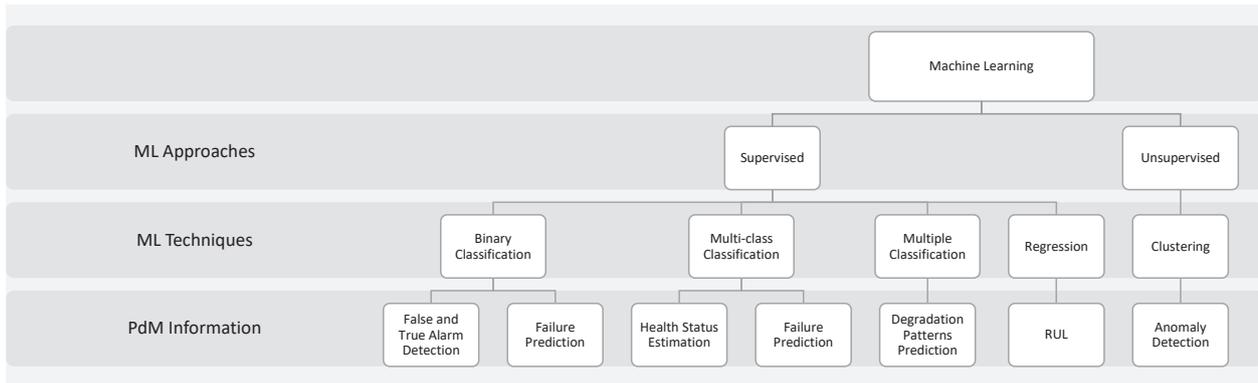


Fig. 2. Machine Learning Techniques for Predictive Maintenance

TABLE V  
MULTIMODAL MACHINE LEARNING (MML) METHODS

Representation	Learning to represent heterogeneous information in a unitary way, easy to be understood and processed by a learning model.
Translation	Mapping the information from one modality to another in a most accurate way.
Alignment	Identifying the inherent relations between sub-components. It also implies dealing with similarity measurements.
Fusion	Joining/combining in a meaningful way the information from different modalities.
Co-learning	Transferring knowledge among modalities: the modality with limited resources can benefit from another with more information.

dependency at the lowest level of features (or raw input unprocessed data), (ii) intermediate-fusion assumes a dependency at a more abstract, semantic level; and (iii) decision-based fusion assumes no dependency at all in the input, but only later at the level of decisions. The above described assumption has the following implications, as argued in [47]: (i) there are no established, standard methods to identify feature dependencies in multiple sensors and modalities; (ii) the technology exists, but there are no standard methods to extract unbiased feature from raw data, and therefore deep learning methods are preferred; (iii) there are basic techniques to handle modality fusion when dealing with missing information; (iv) it is unclear what are the relevant features to be learned, in the sense that a trial-and-error process of feature engineering is employed for the shallow ML algorithms, i.e. Decision Tree, SVM, kNN; and (v) multimodal data fusion best practices i.e. data sets, fusion algorithms, success stories, training and evaluation of results, should be recorded and shared. Moreover, the review of existing proposals for data fusion techniques and frameworks clearly shows that the actual trend for maintenance engineering is cloud maintenance i.e. maintenance-as-a-service, as argued in [2] and [11]. The envisioned platform is seen as a management system of smart services i.e. data-analysis-as-a-service, prognostics-as-a-service or data-as-a-service, that represent better solutions in terms of technology, performance, and costs. The list of challenges continues with: (vi) the absence of a clearly defined generic framework for smart services that standardize the usage of a data fusion pipeline it is clear that in an operational environment more than one data fusion techniques should be applied; (vii) there are no standard

techniques for dealing with temporal and spatial (context) data alignment and synchronization, i.e. the ontology-based proposal of [44] assumes some benefits due to knowledge access, reasoning and re-use of ontology web-standards; and (viii) lack of research studies to analyze the performance of ML algorithms in a cloud environment.

4) *Model Evaluation*: One challenge for evaluating the performance of learning models is represented by the availability of data: when the data set is not large enough to provide sufficient data quantities for the data training validation and test sets, then methods such as k-fold cross validation and bootstrapping are used to simulate new data. K-fold validation is used to split the original dataset into k folds, and run the learning algorithm k times. Another challenge on model evaluation is represented by the skew or imbalanced data. In any maintenance scenario, the minority data class is represented by the abnormal data i.e. the event-faults. Therefore, in the case of fault prediction, the algorithm needs to identify only a small group of data from the overall historical data. Incorrectly predicting a positive class as a negative may lead to a greater cost than the reverse situation, i.e. the problem of the asymmetric cost. Consequently, performance metrics based on the confusion table are used to evaluate how accurate the algorithm is. Boosted methods, such as boosted decision trees are also used as algorithms for solving the imbalanced data problem. Consequently, the time-dependent split of the imbalanced data is useful for avoiding data overfitting on classification models for fault predictions, and on regression models for predicting the RUL. Other problems to be avoided when evaluating a learning model are variance and bias. The

TABLE VI  
MODEL-AGNOSTIC FUSION TYPES

Feature-based (Early Fusion)	Features from all the modalities are concatenated as one long input and trained by a single learner.
Intermediate (Hybrid Fusion)	There is a single learning model which is trained with a preprocessed input from modalities in the fused layer. It is implemented by neural networks and multi-kernel support vector machines algorithms.
Decision-Based (Late Fusion)	Each modality is trained with a different learning model that independently makes a decision. All decisions generated by learning models are later combined based on a fusing schema.

reference [42] addresses these problems suggesting appropriate solutions.

## V. DISCUSSIONS

The present work reviews maintenance approaches with applicability in the industrial environment. The aim is to identify potential sources and ideas for delivering better data analysis tools and techniques for the optimization of the industrial maintenance processes. Past works on maintenance approaches show that maintenance actions are performed by employing various prediction models and modeling techniques. However, the existent literature does not inform us to which extent the new A.I. technology based on ML methods and techniques is influencing and changing the maintenance approaches in the industrial setup. Consequently, we provide an analytical literature review showing first that among all the existent approaches to maintenance, each of them varying in terms of efficiency and complexity, predictive maintenance seems to best fit the needs of a highly competitive industry setup. Next, we consider ML to be a prediction methodology and we show that ML methods enhance industrial maintenance with a critical component of intelligence: prediction. The approach we envision for the optimization of predictive maintenance actions investigates the MML perspective and consequently uses a variety of multimodal ML methods that study both live and historical information, in order to learn prognostics data and to make accurate diagnostics and predictions. Based on the surveyed literature we construct taxonomies that cover the main predictive models and their modeling techniques relative to maintenance goals. We show that among all the prediction models, the data driven, statistical inference based ML approaches are the most suitable to deal with big volumes of heterogeneous data. Their acceptance in the field is mainly due to the fact that prediction is easier than model inference i.e. the ML models are performing tests to check how well a learning model which is trained on a data set is able to predict new data. This allows ML algorithms to easily work with larger volumes of complex data. However, a critical analysis of ML algorithms and of the sensor data sets used for maintenance will directly show that there are no optimal ML models that always outperform all the other. Usually, their efficiency is based on the type of training data distribution. On its turn, multimodality is presented as an efficient ML method of combining data from multiple, diverse modalities and sources. Its main goals are: making better predictions and retrieving new insights from the combined knowledge. A model-agnostic taxonomy of the reviewed multimodal ML fusion methods is presented together with appropriate solu-

tions for optimal usage. In particular, we distinguish among: (i) feature-based fusion or early fusion a basic concatenation of features belonging to different modalities; (ii) intermediate-fusion typical for algorithms implemented by the artificial neuronal networks or by multi-kernel support vector machines; and (iii) decision-based fusion which applies a learning model for each modality independently, and the fusion takes place only at the decision level. Past works present multimodal fusion strategies made in an ad-hoc way, without following some standard implementation lines. We highlight the fact that there is a need for standardized solutions in applying multimodal ML methods for maintenance purposes. Moreover, we show that analyzing only the technical aspect i.e. the multimodal ML perspective, for improving the quality of maintenance is not sufficient. The connections with the economic aspect and the conformity of data science projects with industrial standards like CRISP-DM and Industrial Data Space are relevant. Consequently, we argue that quality of maintenance in an industrial setup can be improved only when in the development of a generalized architecture for maintenance purposes the following aspects are taken into consideration: (i) the technological aspect which recognizes the potential of multimodal ML methods for maintenance purposes; (ii) the business aspect which envisions a structured development of the implementation works starting with the business model's conceptualization, and assuring its conformity with the industry standards; and (iii) the economic aspect which follows the classical optimization concerns relative to maintenance costs. Future works are planned to analyze the usage of multimodal ML methods combined with semantic technologies in a cloud-oriented environment. The goal is to overcome the problem of sensor integration for efficient data analysis. We recognize that the actual trend for maintenance engineering is cloud maintenance. Within this context, the envisioned digital platform is seen as a management system of smart services i.e. prediction-as-a-service and maintenance-as-a-service, with expected benefits in terms of technology, performance and costs.

## REFERENCES

- [1] G. A. Susto, S. Mcloone, S. Pampuri, A. Benghi, and A. Schirru, "Machine Learning for Predictive Maintenance: A Multiple Classifier Approach", *IEEE Transactions on Ind. Inf.* 11(3), 2015, pp. 812-820, <https://doi.org/10.1109/TII.2014.2349359>.
- [2] Z. Liu, M. Norbert, and M. Nezih, The role of Data Fusion in predictive maintenance using Digital Twin, in *AIP Conference Proceedings* 1949(1):02023, 2018, <https://doi.org/10.1063/1.5031520>.
- [3] G. Manco, E. Ritacco, P. Rullo, L. Galluci, W. Astill, D. Kimber, and M. Antoneli, Fault detection and explanation through big data analysis on sensor streams, in *Expert Syst. Appl.* 87, 2017, pp. 141-156, <https://doi.org/10.1016/j.eswa.2017.05.079>.

- [4] G. Niu and H. Li, IETM centered intelligent maintenance system integrating fuzzy semantic inference and data fusion, in *Microelectron. Reliab.* 75, 2017, pp. 197-204, <https://doi.org/10.1016/j.microrel.2017.03.015>.
- [5] T. Widmer, A. Klein, P. Wachter, and S. Meyl, Predicting Material Requirements in the Automotive Industry using Data Mining, in *BIS*, 2019, pp. 582-588.
- [6] Ł. Sobaszek, A. Gola, and E. Kozłowski, Application of survival function in robust scheduling of production jobs, in *FedCSIS 2017*, ACSIS, Vol. 11, 2017, pp. 575-578, <http://dx.doi.org/10.15439/2017F276>.
- [7] Ł. Sobaszek, A. Gola, and E. Kozłowski, Job-shop scheduling with machine breakdown prediction under completion time constraint, in *FedCSIS 2018*, ACSIS, Vol. 15, 2018, pp. 437-440, <http://dx.doi.org/10.15439/2018F83>.
- [8] L. Guo, N. Li, F. Jia, Y. Lei, and J. Lin, A recurrent neural network based health indicator for remaining useful life prediction of bearings, in *Neurocomputing* 240, 2017, pp. 98-109, <https://doi.org/10.1016/j.neucom.2017.02.045>.
- [9] R. Acorsi, R. Manzini, P. Pascarella, M. Patella, and S. Sassi, Data Mining and Machine Learning for Condition-based Maintenance, in *Int. Conf. on Flexible Automation and Intelligent Manufacturing* 11, 2017, pp. 1153-1161, <https://doi.org/10.1016/j.promfg.2017.07.239>.
- [10] M. Safizadeh and S. Latifi, Using multisensory data fusion for vibration fault diagnosis of rolling element bearings by accelerometer and load cell, in *Inf. Fusion* 18, 2014, pp. 1-8, <https://doi.org/10.1016/j.inffus.2013.10.002>.
- [11] B. Schmidt, U. Sandberg and, L. Wang, Next generation condition based Predictive Maintenance, in *Methods* 13306, 2014, pp. 4-11.
- [12] M. Schenk, Instandhaltung technischer Systeme, 2010.
- [13] B. Otto, S. Auer, J. Cirullies, J. Jürjens, N. Menz, J. Schon, and S. Wenzel, Industrial Data Space Digital sovereignty over data, in *Fraunhofer Gesellschaft zur Förderung der angewandten Forschung*, 2016.
- [14] DIN EN-13306. DIN Standards Publication Maintenance Begriffe der Instandhaltung/Maintenance terminology, 2010.
- [15] DIN EN-31051. DIN Standards Publication Maintenance Grundlage der Instandhaltung/Fundamentals of Maintenance, 2012.
- [16] A.R. Hevner, S.T. March, J. Park, and S. Ram, Design science in information system research, *MIS Q.* 28(1), 2004, pp. 75-105.
- [17] B. J. Oates, Researching Information Systems and Computing, *Sage Publications Ltd.*, 2006.
- [18] K. Peffers, T. Tuunanen, M. Rothenberger and S. Chatterjee, A Design Science Research Methodology for Information Systems Research, in *J. Manage. Inf. Syst.* 24(3), 2007, pp. 45-77, <https://doi.org/10.2753/MIS0742-122240302>.
- [19] C. Bunks, D. McCarthy, and T. Al-Ani, Condition-based Maintenance of machines using hidden Markov Models, in *NAMRC* 32, 2004, pp. 597-612, <https://doi.org/10.1006/mssp.2000.1309>.
- [20] P. Deuzkiewicz and S. Radkowski, On-line condition monitoring of a power transmission unit of a rail vehicle, in *Mechanical Systems and Signal Processing* 17(6), 2003, pp. 1321-1334, <https://doi.org/10.1006/mssp.2002.1578>.
- [21] Y. Hao, J. Sun, G. Yang and J. Bai, The Application of Support Vector Machines to Gas Turbines Performance Diagnosis, in *Chinese Journal of Aeronautics* 18 (1), 2005, pp. 15-19, [https://doi.org/10.1016/S1000-9361\(11\)60276-8](https://doi.org/10.1016/S1000-9361(11)60276-8).
- [22] P. Baraldi, E. Zio and F. di Maio, Unsupervised Clustering for Fault Diagnostics in Nuclear Power Plants Components, in *Int. Journal of Comp. Intelligent Systems* 6(4), 2014, pp. 764-777, <https://doi.org/10.1080/18756891.2013.804145>.
- [23] A. Alexandru, Using Expert Systems for Fault Detection and Diagnosis in *Industrial Applications*, 1998.
- [24] A. Krishnakumari, A. Elayaperumal, M. Saravanan, and C. Arvindan, Fault diagnostics of spur gear using decision tree and fuzzy classifier, in *Int. J. Adv. Manuf. Technol.* 89 (9-12), 2017, pp. 3487-3494, <https://doi.org/10.1007/s00170-016-9307-8>.
- [25] V.H. Jaramillo, J.R. Ottewill, R. Dudek, D. Lepiarczyk, and P. Pawlik, Condition monitoring of distributed systems using two-stage Bayesian inference data fusion, in *Mech. Syst. Signal Process.* 87, 2017, pp. 91-110, <https://doi.org/10.1016/j.ymssp.2016.10.004>.
- [26] C. Liu, Y. Li, G. Zhou, and W. Shen, A sensor fusion and support vector machine based approach for recognition of complex machining conditions, in *Journal of Intelligent Manufacturing*, 2016, pp. 1-14, <https://doi.org/10.1007/s10845-016-1209-y>.
- [27] A. Diez, N.L.D. Khoa, M.M. Alamdari, Y. Wang, F. Chen, and P. Runcie, A clustering approach for structural health monitoring on bridges, in *J. Civil Struct. Health Monitoring* 6 (3), 2016, pp. 429-445.
- [28] C. Li, R.-V. Sánchez, G. Zurita, M. Cerrada, and D. Cabrera, Fault diagnosis for rotating machinery using vibration measurement deep statistical feature learning, in *Sensors* 16 (6): 895, 2016, pp. 1-19.
- [29] Q. (C.) Liu and H.P. (B.) Wang, A case study on multisensory data fusion for imbalanced diagnosis of rotating machinery, in *AI EDAM* 15(3), 2001, pp. 203-210.
- [30] A. Xenakis, A. Karageorgos, E. Lallas, A.E. Chis, and H. Gonzalez-Velez, Towards Distributed IoT/Cloud based Fault Detection and Maintenance in Industrial Automation, in *EDI40*, 2019, pp. 683-690, <https://doi.org/10.1016/j.procs.2019.04.091>.
- [31] A. Mosallam, K. Medjaher, and N. Zerhouni, Data-driven prognostic method based on Bayesian approaches for direct remaining useful life prediction, in *J. Intell. Manuf.* 27 (5), 2016, pp. 1037-1048, <https://doi.org/10.1007/s10845-014-0933-4>.
- [32] E. F. Alsina, M. Chica, K. Trawinski, and A. Regattieri, On the use of Machine Learning methods to predict component reliability from data-driven industrial case studies, in *Int. J. Adv. Manufacturing Technology*, (94), 2018, pp. 2419-2433, <https://doi.org/10.1007/s00170-017-1039-x>.
- [33] L. Cristaldi, G. Leone, R. Ottoboni, S. Subbiah, and S. Turin, A comparative study on data-driven prognostic approaches using fleet knowledge, in *IEEE International Conference on Instrumentation and Measurement Technology (I2MTC)*, 2016, pp. 1-6, <https://doi.org/10.1109/I2MTC.2016.7520371>.
- [34] C. F. Baban, M. Baban, and M.D. Suteu, Using a fuzzy logic approach for the predictive maintenance of textile machines, in *J. Intell. Fuzzy Syst.* 30 (2), 2016, pp. 999-1006, <https://doi.org/10.3233/IFS-151822>.
- [35] W. Cui, Z. Lu, C. Li, and X. Han, A proactive approach to solve integrated production scheduling and maintenance planning problem in flow shops, in *Comput. Ind. Eng.* 115, 2018, pp. 342-353, <https://doi.org/10.1016/j.cie.2017.11.020>.
- [36] T. Baltrusaitis, C. Ahuja, and L. Morency, Multimodal Machine Learning: A Survey and Taxonomy, in *IEEE transactions on pattern analysis and machine intelligence*, 2017, pp. 423-443, <https://doi.org/10.1109/TPAMI.2018.2798607>.
- [37] E. Alpaydin, Classifying multimodal data" in *The Handbook of Multimodal-Multisensor Interfaces*, Ed. Sharon Oviatt, Björn Schuller, Philip R. Cohen, Daniel Sonntag, Geranimos Potamianos, and Antonio Krüger, in Association for Computing Machinery and Morgan & Claypool, NY, 2018, pp. 49-69, <https://doi.org/10.1145/3107990.3107994>.
- [38] B. Khaleghi, F. Karray, A. Khamis, and S. N. Razavi, Multisensor Data Fusion: A review of the State-of-the-Art, in *Information Fusion* 14, 2013, pp. 28-44, <https://doi.org/10.1016/j.inffus.2011.08.001>.
- [39] Y. Bengio, A. Courville, and P. Vincent, Representation learning: a review and new perspectives, Technical report. U Montreal, 35(8), pp. 1798-1828, 2013, <https://doi.org/10.1109/TPAMI.2013.50>.
- [40] N. Srivastava and R. Salakhutdinov, Multimodal learning with Multimodal Boltzmann Machines, in *Advances in Neural Information Processing Systems*, 2012, pp. 2222-2230.
- [41] A. Zheng, Evaluating machine Learning Models A Beginners Guide to Key Concepts and Pitfalls. O'Reilly Media, 2015, ISBN 978-1-491-93246-9.
- [42] A. Ng, Machine Learning, Online Course offered by Stanford University, 2010.
- [43] B. Schmidt, W. Lihui, and D. Galar, Semantic Framework for PdM in a cloud environment, in *CIRP ICME* 62, 2017, pp. 582-588, <https://doi.org/10.1016/j.procir.2016.06.047>.
- [44] H. Seidgar, M. Zandieh, and I. Mahdavi, An efficient meta-heuristic algorithm for scheduling a two-stage assembly flow shop problem with preventive maintenance activities and reliability approach, in *Int. J. Ind. Syst. Eng.* 26(1), 2017, pp. 16-41, <https://doi.org/10.1504/IJISE.2017.083180>.
- [45] A. Diez-Olivan, J. del Ser, D. Galar, and B. Sierra, Data fusion and machine learning for industrial prognosis: Trends and perspectives towards Industry 4.0, in *Int. J. Inf. Fusion* 50, 2019, pp. 92-111, <https://doi.org/10.1016/j.inffus.2018.10.005>.
- [46] P.H. Foo and G.W. Ng, High-level Information Fusion: An Overview, in *Journal of Advances in Information Fusion*, 8(1), 2013, pp. 33-72, <https://doi.org/10.1.1.360.6651>.
- [47] C.-A. Chou, X. Jin, A. Müller, and S. Ostadabbas, (MMDF) Multimodal Data Fusion Workshop Report, 2018.

## Visual Rule Editor for E-Guide Gamification Web Platform

Artur Kulpa  
University of Szczecin,  
Faculty of Economics and  
Management,  
ul. Mickiewicza 64, 71-101  
Szczecin, Poland  
Email: artur.kulpa@usz.edu.pl

Jakub Swacha  
University of Szczecin,  
Faculty of Economics and  
Management,  
ul. Mickiewicza 64, 71-101  
Szczecin, Poland  
Email: jakub.swacha@usz.edu.pl

Karolina Muszyńska  
University of Szczecin,  
Faculty of Economics and  
Management,  
ul. Mickiewicza 64, 71-101  
Szczecin, Poland; Email:  
karolina.muszynska@usz.edu.pl

**Abstract**—Gamification is applied in different information systems to motivate the users and make their experience with the system richer and more engaging. Gamification employed in e-guides aims at enhancing the process of visiting a tourist attraction. Even though each tourist attraction is unique and requires an individual gamification scheme, similarities in the components and procedures used to develop such schemes led to the development of a generic e-guide gamification framework. One of its main principles is to store the gamification rules as a content separate from the engine to process them. This way, the rules can be easily edited by subject matter experts. This paper describes a visual rule editor developed to facilitate this process.

### I. INTRODUCTION

GAMIFICATION, understood as “the use of game-design and game psychology in non-game settings to engage the target audience and motivate specific behaviors” [1], has been considered as “one of the significant new trends in the development of services and applications in the software industry” [2]. It can be applied to information systems of various character [3]. One of these are multimedia visitor guidance systems, better known as e-guides, which may employ gamification to enhance the tourist attraction visiting process [4].

Tourist attractions may significantly differ in their character, and e-guides may as well differ in their form and functionality. For this reason, there cannot be a fixed scheme for e-guide gamification. Rather, individual gamification schemes have to be designed for respective tourist attractions with specific characteristics of their own. There is still significant similarity among components and procedures used to develop gamification schemes for different tourist attractions. Motivated by this observation, a generic e-guide gamification framework has been proposed, aiming to standardize architectural and design solutions so

that it could be easier to implement gamification to e-guides, reuse gamification layer among different e-guides, and maintain and update gamified e-guides [5].

An indispensable ingredient of any gamification scheme are the rules linking users’ actions to game-inspired consequences, such as specific rewards or feedback [6]. While they could be developed as a part of e-guide software, the framework mentioned above proposes not only the separation of the core e-guide functionality and content from the gamification engine and rules, but also the separation of gamification rules and the engine to process them, so that the rules could be treated as a kind of content, separately edited and transferred among e-guides [5].

For this purpose, a common notation of gamification rules for e-guides has to be used, and such notation has been proposed [7]. Although, the textual notation it uses was designed for readability, in the course of implementation of the BalticMuseums: Love IT! project [8], it was found that the tourist attraction personnel who was responsible for devising the e-guided tours featuring gamification, and who had no programming experience found it difficult to use.

In this paper, we describe a solution for this problem in a form of a dedicated visual rule editor, so that the prior knowledge of the notation syntax, keywords, and appropriate parameters is no longer required from the gamification designer. This approach follows the example of general-purpose task automation services where the provision of visual rule editor is considered as an architectural requirement [9, p. 13].

The structure of the paper is as follows: section II gives a short glimpse of prior work on this topic. In section III, we present the user interface of the visual rule editor, so that the reader could see how it can be used to define gamification rules. Section IV provides necessary technical information about the implementation of the visual rule editor. In section V, the results of its evaluation are reported. The final section concludes.

---

The presented rule editor was developed as a part of the BalticMuseums: Love IT! project and part-financed from the European Regional Development Fund within the Interreg South Baltic Programme.

This paper was developed with a financial support from a project financed within the framework of the program of the Minister of Science and Higher Education under the name "Regional Excellence Initiative" in the years 2019-2022; project number 001/RID/2018/19; the amount of financing PLN 10,684,000.00.

## II. RELATED WORK

The idea of replacing text statements with a form to be filled in can be traced back to QBE database query language released commercially already in 1978 [10]. In its contemporary implementations (such as in Microsoft Access 2019), not only the query building form has a predefined structure (so that there is no need to know the syntax), but also the available options are simply checked or unchecked (rather than typed in) and wherever the set of possible parameter values is predefined, they can be chosen from a list rather than input as text (which has to be done with other parameter values).

Although some research results show that trained and experienced users prefer textual notation than forms (see e.g. [11]), for a first-time or occasional user, the advantage of using a form over writing text in a language unknown to him/her is obvious. This is a reason for which a number of similar query languages were developed, including [12]:

- Aggregates-by-Example, Summary-Table-by-Example, and Query-by-Statistical-Relational-Table designed for querying statistical and scientific databases;
- Time-by-Example designed for querying historical databases;
- Generalized-Query-by-Example designed for querying relational, network, and hierarchical databases;
- Picquery and Query-by-Pictorial-Example designed for querying image databases.

As gamification rules, like database queries, also have a well-defined context and structure, they are suitable for editing using forms. This has been proven by well-known enterprise gamification platforms, such as Bunchball Nitro [13] or Gametize [14], which use forms to specify the rules.

In the following two sections, we will show how this approach has been implemented to the e-guide gamification rule specification.

## III. VISUAL RULE EDITOR

The described visual rule editor forms a part of the e-guide content management system which in turn is a part of the e-guide gamification web platform developed within the BalticMuseums: Love IT! international project [8]. The purpose of the editor is to facilitate gamification designers to add, view, modify and delete gamification rules which are to be applied by the e-guide gamification web service [15], itself called by e-guide client applications running on tourist attraction visitors' mobile devices.

The visual rule editor features a list of defined rules along with the search and pagination mechanisms, as shown in Fig. 1. The editor follows the principle of a clear division of rule specification into three parts: the name (identifying a rule), conditions (required to trigger a rule) and results (caused by a rule) parts as presented in [5]. This allows for easy specification of rules having multiple conditions and results. Both the conditions and results regarding a given rule are displayed next to its name in a readable form and the standard functionalities of adding, viewing, modifying and deleting them are provided.

Adding and modifying rules is done using the "Edit rule" form presented in Fig. 2. Similar forms are used to add or edit a condition or a result for a selected rule.

Fig. 2 Rule edit form

No.	Name	Description	Results	Active	Conditions
1.	Baltic Adventure	Towards a new adventure.	<ul style="list-style-type: none"> <li>x text: Welcome</li> <li>x points: 3</li> <li>x badge: Captain Nemo:Captain Nemo</li> </ul> + Add result	✓	<ul style="list-style-type: none"> <li>x variable [Score] &gt; value [320]</li> <li>x location in [Flatfish and gobies, Giant clam and dartfish]</li> <li>x item in [Puzzle to solve]</li> <li>x date &gt;= [2019-05-01]</li> <li>x date &lt;= [2019-05-31]</li> <li>x time &gt;= [08:00:00]</li> <li>x time &lt;= [12:00:00]</li> </ul> + Add condition

Fig. 1 Rules definition module in the e-guide gamification system

Figure 3 presents a condition edit form with a *Var\_Value* type condition, which makes it possible to specify: the name of the variable that is to occur in the description of the event that triggers the rule, the comparison operator and the threshold value at which the condition is met. For the user’s convenience, other conditions specified for the rule whose condition is being edited are shown above the edit text boxes. The available comparison operators include those for comparing two values (=, <>, >, >=, <, <=) and checking the existence of a value in a set (*in*, *not in*).

Fig. 3 Condition edit form

Figure 4 presents a result edit form with an example of a *Badge* type result. The other two result types currently available are *Points* and *Text* (to be displayed to the visitor). The list of result types is to be extended in consideration of the detailed results of the evaluation survey (see section V). As with editing the conditions, for the user’s convenience, other results specified for the rule whose condition is being edited are shown above the edit text boxes.

Fig. 4 Result edit form

#### IV. DATA MODEL FOR REPRESENTATION OF RULES

The data model for the representation of rules of the gamification service is based on classes implemented in the Django framework (Fig. 5). During implementation, this model is translated into SQL, and then a SQL script is executed in the PostgreSQL database management system, which results in the creation of the respective tables. The

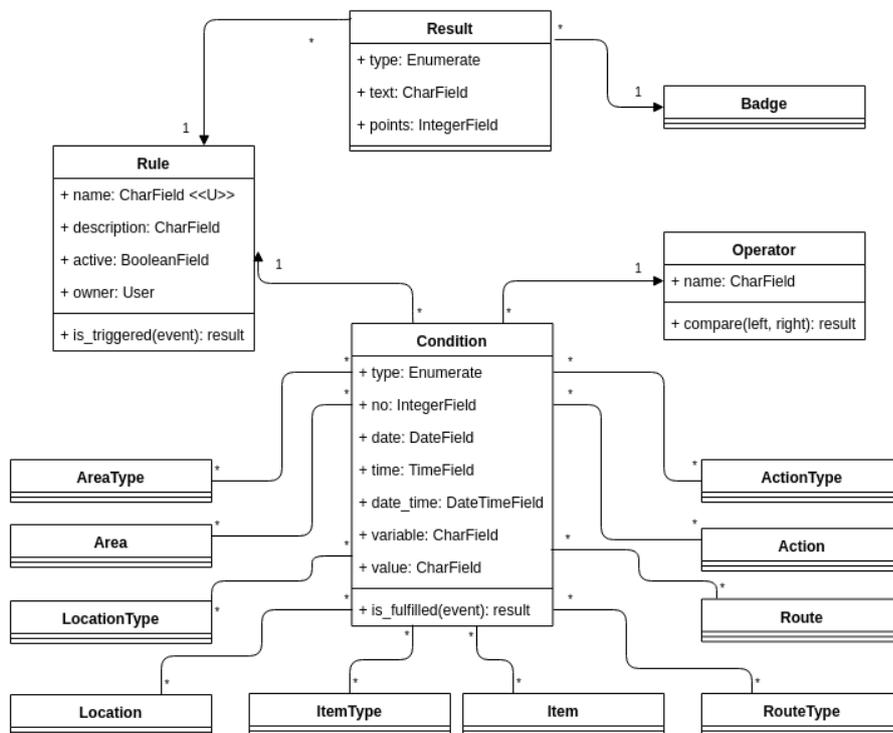


Fig. 5 Data model for representation of rules

main class of the model is the Rule class which defines objects storing basic information about each rule – its *name* and *description*. The content of the *name* field has to be unique (no two rules may have the same name). An additional *active* field has been introduced to indicate whether, for a given rule, the data stored in objects of associated classes (see below) can be modified or not, as well as whether the rule can be executed by the e-guide gamification web service. Each rule is assigned to the user who created it. This way, only an owner of the rule can see and modify it.

The *is\_triggered* method defined within the Rule class checks if the rule has been triggered as a result of an event (whose description is the input parameter) and if positive, returns the results assigned to the rule.

The Result class contains a field specifying the result type. Currently, there are three available result types: Text, Points, and Badge. Depending on the result type, the appropriate field is filled in. In the case of Text type, the *text* field is set to the result message to the visitor. In the case of Points type, the *points* integer field is set to the number of points to be received by the visitor. In the case of Badge type, an object from the Badge class specifying the badge to be awarded to the visitor is assigned to the Result class object.

The Condition class is meant to map any conditions associated with a given rule. One rule can have many conditions and a certain condition can be associated with only one rule. The *type* field specifies the object of the left side of the condition. There are several entities in the e-guide gamification system that can be used to form the right side of the condition. They can be seen on the data model associated with the Condition class, i.e.: *AreaType*, *Area*, *LocationType*, *Location*, *ItemType*, *Item*, *RouteType*, *Route*, *ActionType* and *Action*. Other available condition types include *Date*, *Time*, *Date\_Time* and *Var\_Value*. For the *Var\_Value* type, both the variable name and its threshold value must be specified.

The *is\_fulfilled* method, defined within the Condition class, specifies if a certain condition has been fulfilled, based on an event description (the input parameter).

Each object of the Condition class is associated with an object of the Operator class which defines the way of comparing both sides of the condition. The *compare* method of the Operator class is used to perform the comparison.

## V. EVALUATION

The preliminary evaluation of the described visual rule editor was based on a qualitative survey among the representatives of tourist attractions participating in the BalticMuseums: Love IT! project [8] who are in progress of development of their respective gamified e-guides and to whom the tool was presented earlier. The three rule editor

evaluation questions constituted a part of a larger survey aimed at defining future work directions and possible technical improvements, therefore the answers could be considered as objective. Out of the five tourist attractions participating in the BalticMuseums: Love IT! project, answers from four (Gdynia Aquarium, Experiment Science Center, Malmö Museums and NaturBornholm) were received (six answers in total as there were two organizations with two representatives from each of them participating in the survey). Note that though the number of the tourist attractions involved in the survey is low, they differ significantly in their character, which makes their answers far more representative than if they all belonged to a single category of attractions.

Answering to the first evaluation question, all the six surveyed representatives agreed that a form-based gamification rule editor is better than writing rules as text (using a dedicated rule definition language). This confirms the original observations that led to the development of the editor.

The second question asked whether there were rules that the tourist attraction representatives invented but did not manage to define using the editor. Half of the surveyed (three of six representatives from two of the four involved attractions) answered positively, with one respondent pointing to her (or her team's) possible lack of skills in using the tool, and the two other (from one attraction) suggesting that the rule definition form has to be extended.

In the third question, the respondents were asked if they encountered other problems with defining rules. Again, half of the surveyed (three of six representatives from two of the four involved attractions) answered positively. This time, all of them pointed to the lack of a good manual as a reason for those problems.

## VI. CONCLUSION

A key point in the implementation of gamification is the specification of rules governing the system. While such rules may be embedded in the gamification software, in many circumstances it makes a lot of sense to have them defined separately so that a change in gamification rules would not require a change in the software handling them.

This is especially true in the case of e-guide gamification where the rules should make use of the specificity of a tourist attraction, which strengthens the role of subject matter experts (such as guides and educators) rather than gamification experts. Another reason for such an approach to e-guide gamification is due to quickly changing content featured in many tourist attractions (such as those having seasonal exhibits or short-living fauna species on display).

While e-guide gamification rules can be effectively represented in textual notation [7], their editing by subject

matter experts (rather than IT professionals) could be much simpler using predefined forms.

The visual rule editor described in this paper was designed following this approach. All the surveyed representatives of the tourist attractions introduced to it unanimously agreed on their preference for the form-based rather than textual rule specification. On the other hand, the survey revealed that the decision to make the editor simple by limiting the possible scope of defined rules (as compared to the textual notation) resulted in some of the tourist attractions representatives being unable to define sophisticated rules they had invented. Moreover, even though the operation of the editor is intuitive, the users encountered some problems which could be addressed by providing a better user manual.

The identified drawbacks will be addressed in the next version of the rule editor, whose development constitutes the nearest future work, after which another evaluation survey is planned, having an extended scope and involving more tourist attractions.

#### REFERENCES

- [1] A. Marczewski, "Defining gamification – what do people really think?," <http://www.gamified.uk/2014/04/16/defining-gamification-people-really-think/>, last accessed 31.5.2018.
- [2] J. Kasurinen and A. Knutas, "Publication trends in gamification: A systematic mapping study," *Computer Science Review*, vol. 27, 2018, pp. 33–44. doi:10.1016/j.cosrev.2017.10.003.
- [3] J. Swacha, „Gamification in Enterprise Information Systems: What, Why and How,” in: *Proceedings of the Federated Conference on Computer Science and Information Systems, Annals of Computer Science and Information Systems*, vol. 8, 2016, pp. 1229–1233. doi: 10.15439/2016F460.
- [4] J. Swacha and R. Ittermann, "Enhancing the tourist attraction visiting process with gamification: key concepts," *Engineering Management in Production and Services*, vol. 9, no. 4, pp. 59–66, 2017. doi:10.1515/emj-2017-0031.
- [5] J. Swacha and K. Muszynska, "Towards a Generic eGuide Gamification Framework for Tourist Attractions", in: *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts, CHI PLAY 2018*, Melbourne, 2018, pp. 619–625. doi: 10.1145/3270316.3271535.
- [6] J. Swacha, „Architecture of a dispersed gamification system for tourist attractions," *Information*, vol. 10, no. 1, 2019, art. 33. doi: 10.3390/info10010033.
- [7] J. Swacha, "Representation of Events and Rules in Gamification Systems," *Procedia Computer Science*, vol. 126, 2018, pp. 2040–2049. doi: 10.1016/j.procs.2018.07.248.
- [8] BalticMuseums: Love IT! Project website. <http://www.balticmuseums.info>, last accessed 15.5.2019.
- [9] M.C. Barrios, "A Personal Agent Architecture for Task Automation in the Web of Data. Bringing Intelligence to Everyday Tasks," PhD Thesis, Universidad Politécnica de Madrid, Madrid, 2016.
- [10] M. M. Zloof, "Query-by-Example: A data base language," *IBM Systems Journal*, vol. 16, no. 4, pp. 324–343, 1977. doi: 10.1147/sj.164.0324.
- [11] J. M. Boyle, K. F. Bury, and R. J. Evey, "Two Studies Evaluating Learning and Use of QBE and SQL," *Proceedings of the Human Factors Society Annual Meeting*, vol. 27, no. 7, pp. 663–667, 1983. doi: 10.1177/154193128302700732.
- [12] G. Özsoyoglu and H. Wang, "Example-based graphical database query languages," *Computer*, vol. 26, pp. 25–38, 1993. doi: 10.1109/2.211893.
- [13] Nitro: The Enterprise Engagement Platform Powered by Gamification, <https://www.bunchball.com/products/nitro-platform>, last accessed 15.5.2019.
- [14] Community Engagement, Gametized, <https://gametize.com/index>, last accessed 15.5.2019.
- [15] J. Swacha and A. Kulpa, "A Cloud-Based Service for Gamification of eGuides," in 2018 6th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW), Barcelona, 2018, pp. 220–224. doi: 10.1109/W-FiCloud.2018.00042.



# A Design and Experiment of Automation Management System for Platform as a Service

Alalaa Tashkandi

Saudi Aramco, Information Technology, Al-Midra Tower, Dhahran 31311, Saudi Arabia  
Email: alaa.tashkandi@aramco.com

**Abstract**—Security [11] and quality [4] of cloud computing (CC) services represent significant factors that affect the adoption by consumers. Platform as a Service (PaaS) is one of CC service models [14]. Management of database systems, middleware and application runtime environments is automated in PaaS [2]. PaaS automation management issues and requirements were collected in three rounds from information technology experts using Delphi technique. In this paper, PaaS automation quality and security management system (MS) layered model is proposed and validated. The aim of the MS is enabling PaaS model for mission critical platforms. Validation of the MS was based on experiment in a private cloud for an organization undergoing a transformation toward PaaS computing.

## I. INTRODUCTION

SERVICE interruption caused by Platform as a Service (PaaS) automation failure or security incidents may lead to reputation damage or reimbursement cost. PaaS management system (MS) should provide integrated proactive and reactive control for quality and security events.

The MS model was achieved based on analysis of experts' inputs using Delphi technique in a Cloud Service Provider (CSP) organization providing mission critical services. The model enforces secure rollout, update and monitoring of automation artifacts. Additionally, quality assurance (QA) process is enforced. After three rounds of communications, verifications and feedbacks, PaaS MS model was finalized. Observations based on MS experiment are summarized in Analysis and Discussion section of this paper.

In this research, we have collected the adoption factors and requirements of PaaS for mission critical platforms to drive a novel management model using Delphi approach. The model was experimented and validated in a private cloud environment.

## II. RELATED WORK

Applications in PaaS context can be classified based on computation model, resource utilization type, resource utilization variability or interactivity level. Examples of PaaS offering include Google App Engine and Amazon Web Services Lambda [2]. PaaS is one of three service models in

Cloud Computing (CC). In this model, customer does not have control over the infrastructure layer while he has control over the application and its configuration [14]. Middleware, application runtime environment and database systems are examples of components in PaaS model [3], [5], [10], [13]. PaaS provides self-deployment, monitoring and lifecycle management of applications [2], [4]. PaaS offers include middleware services, like Database as a Service (DbaaS), and reusable middleware components. Middleware components include application runtime environments that can be used as part of an application platform [5], [8].

PaaS provides flexibility and agility to developer team. By automating the deployment of application platforms and their components, different setup and configuration can be realized with minimal efforts and time. Interactions between developer team and operation team are minimized through automation [3], [7]. Infrastructure is abstracted by PaaS [5].

Private cloud and multicloud are now the convention of organizations that need to protect their information technology investment while seeking new technologies and opportunities. Use of resources in multicloud can be sequential or parallel. Support of multicloud minimizes vendor lock-in issue associated with the Cloud [2], [7], [12].

Two sets of users represent PaaS consumers. The first set represents software developers in organizations. The other set represents SaaS CSPs who needs to focus on Software development and services quality [7].

### A. PaaS Challenges

The use of CC in general provides speed and flexibility to organization. On the other hand, cloud raises challenges in terms of quality and security [4], [11]. When PaaS is evaluated by organization for production and business operation, security and QA are two significant adoption factors. Service Level Agreement (SLA) is one of the solutions that was studied in several papers [2], [16]. Service level quality is measured based on Service Level Objectives (SLOs). Service Level Agreement (SLA) grants SLOs based on contractual agreement between the service provider or a broker and the customer [2]. Sequence of events and logs of security control and QA must be forensically sound and reliable in case of a security incident or PaaS failure [1]. Under PaaS service model, wide varieties of technologies exist [4]. Spe-

This work was sponsored by Computer Operations Department, Aramco.

cific applications in PaaS are based on distributed computing. Managing the dependency between distributed components is complex [10].

### B. PaaS Automation

Application deployment automation is enabler for PaaS [6]. Automation performance can be evaluated qualitatively or quantitatively [5]. Operation automation approaches were classified into infrastructure management, plan-based configuration management, image-based configuration management, model-based management and platform centric management automation [9]. Automation of middleware and application deployment should be decoupled from Infrastructure layer. Encapsulating applications in virtual images reduces the flexibility of updating applications over time [6].

Automation standardizations; such as Topology and Orchestration Specification for Cloud Applications (TOSCA); are emerging. TOSCA has relatively small community. Released automation artifacts are limited [9], [15], [17].

## III. DATA COLLECTION

Standards in PaaS are emerging. Security and quality control processes in this context are under development. In this research project, efforts were devoted to build a MS for PaaS that is suitable for organization mission critical operations.

### A. Research Methodology

Delphi approach was followed to collect data and feedback from Information Technology experts within private cloud provider organization. The first round was started after building PaaS automation system without quality and security control in a dedicated computing lab environment. Several Automation scenarios were tested to deploy Database and

application servers, and to manage their lifecycle in PaaS context. Feedbacks from IT security team, datacenter infrastructure team and middleware operation team were collected.

Based on the feedback from the experts, the system was customized with control measures to control the execution of operating system (OS) privileged activities. The measures comprised providing central management of executing privileged commands and logging the activities on central server.

The second round of data collection was started after experimenting the control measures highlighted above. The teams in the previous round were approached to provide their feedbacks after implementing the updates and testing PaaS automation within the same organization.

The third round of data collection was triggered after summarizing experts' feedback and sharing the results of the second round with all participants. The proposed system in this paper was designed and built to address security and quality requirements. In this round, the design was shared with experts for their feedback. Based on their feedback, enhancements were added and final green light was received from experts to build the system for production operation.

### B. Data

Critics were received during the first and second round. Table I provides PaaS automation critics summary. The proposed design in the next section was developed based on the critics and the requirements gathered during the first and second rounds of Delphi process. The design was drafted during the third round of Delphi process. Minor updates were added to the system design in this round. Final design was validated through experiment.

TABLE I.  
SUMMARY OF CRITICS AND EXPERIMENT FINDINGS

#	Critic	Experiment Findings
1	Automation of application platforms is not reliable. Automation quality is low and cannot be used for mission critical systems.	Quality Assurance (QA) should be integrated in the process of PaaS automation development. The design includes lab environments to simulate scenarios and minimize the likelihood of failure. Control measures are added in the system to prevent direct update in production.
2	If automation is doing everything, labor will lose the technical skills and know how.	With new technologies, a shift in labor skills is required. The experiment findings related to this critic are summarized in Analysis and Discussion section.
3	The cost of building PaaS and the required automation is high.	Financial Return on Investment analysis is required to calculate cost against expected return. High level analysis is provided in Analysis and Discussion section.
4	Update of policies and procedures of managing and operating application platforms is required.	This is confirmed in the experiment. The update was mandatory to accommodate PaaS objectives.
5	Automating the deployment of applications involves multiple functional groups in Information Technology department.	The proposed design in this paper addresses Segregation of Duties (SoD) requirements. SoD is leveraged in the design to increase the level of automation quality and security.
6	Automation management server is empowered with privileged accounts and connected to applications and infrastructure components across the organization.	The design provides control for the privileged accounts. The power of PaaS production Orchestrator (PO) was contained by limiting the set of allowed operations. The system securely manages the distribution of automation artifacts. Unauthorized updates are detected by dual integrity control subsystem.
7	Specific experts expressed their ability to achieve the same results with traditional scripting approaches done at OS level.	Acceptable level of quality and security to implement mission critical PaaS is achieved in this experiment. To validate the claim, different approaches and designs should be experimented and compared objectively in future work.
8	Application platforms are updated regularly with new features, security and bug fixes. There is an overhead of keeping track of changes and updating automation.	A balance should be established between the value of update and automation development cost. Image based and plan based automation approaches were experimented. After the deployment of an older version, plans are executed to roll forward the platform to the required release.

#### IV. PROPOSED CONTROL DESIGN

In this section, quality and security management design for PaaS is provided. The design is divided into five layers based on roles and responsibilities to achieve the required quality and security control.

##### A. Layered Management System

The MS consists of Security Audit Layer, Infrastructure as a Service (IaaS) Management Layer, PaaS Management Layer, PaaS Automation Developer Layer and PaaS Consumer Layer. Segregation of Duties (SoD) concept is implemented to maximize security and quality control in mission critical systems without losing the opportunity of automating their deployment and routine operational activities. The five layers are reflected in Fig. 1.

Security Audit Layer includes Privileged Activities Management, Security Policies and Security Events systems. Security policies contain Delegated Privileged Commands (DPC)s which give delegate groups the authority to execute privileged commands. IaaS Privileged commands are executed in a controlled approach. First of all, IaaS administrator assess the impact of delegating and automating the command or infrastructure operation. Once assessed and approved, Security Analyst registers them and assigns them to a delegate group in a security policy within Security Audit Layer. Thereby, the infrastructure operation is utilized at PaaS layers as a self-service. Privileged Activities Management server enables a secure and central management of privileged commands. Security Events database logs the usage of DPCs remotely.

IaaS Management Layer is where infrastructure resources of the cloud are managed. Infrastructure resources include hypervisors, bare metal hosts, storage systems and network components. IaaS administrator manages this layer. His activities as a super user are controlled by Privileged Activities Management server. IaaS resources deployment and management are orchestrated by central IaaS orchestrator tool.

The third layer is PaaS Management Layer. This layer consists of PaaS Development Orchestrator (DO), PaaS Production Orchestrator (PO) and shared storage called Platform Images (PI). PaaS administrator manages these components to automate the deployment and management of PaaS applications. He also manages containerization platforms in contrast to hypervisors which are managed by IaaS administrator. PaaS DO consists of an application server and OS level Development Area. The application server is used to model PaaS automation processes by PaaS developer. It is also integrated with lab resources to conduct QA tests by PaaS developer in the next layer. Development Area is used to build OS level artifacts. The Orchestrator orchestrates PaaS automation in labs and provides pooling capability to logically isolate lab resources based on projects. Automation is developed and simulated in isolated environment before deploying it in production to verify security and quality. PaaS PO is generally similar to the DO. PO is connected to consumer assigned tenants and resources.

PI storage in PaaS Management Layer consists of Development Engine, Production Engine, Distributor, Repositories, Images and Shipment areas. PI storage is mounted in all PaaS resources. Development and Production Engines store automation artifacts. In order to release new automation artifacts from PaaS Development Area to Development Engine, Delegated Privileged Commands are used. PaaS Developer is delegated to update Development Engine using pre-approved delegated commands. These delegated commands verify integrity and analyze automation run time requirements. In addition, release event and content of developed artifacts are reflected in the Security Events system. The deployment to Production Engine is also protected. Only PaaS administrator is delegated to deploy through DPCs. This is enforced to verify the quality of automation before deploying them for production use.

The fourth Layer of the MS is PaaS Automation Developer which controls the development and QA of automation associated with PaaS. The control is enforced by system design. Developer cannot update a production artifact directly. Artifact represents automation logical unit. PaaS automation developers are PaaS automation experts who understand the requirement to deploy and manage applications' platforms.

The fifth layer is PaaS Consumer Layer. PaaS consumer can be SaaS service provider or Organization software developer [7]. PaaS consumer consumes PaaS automated services on his assigned resources through PO.

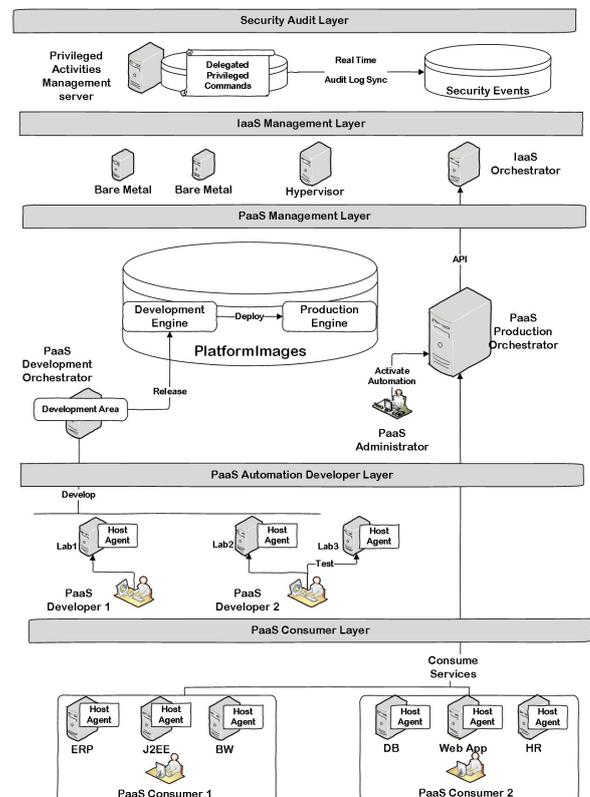


Fig 1. PaaS Management System

## V. EXPERIMENT

The design was internally validated through experiment. It was implemented in a private CC data center.

### A. PaaS Layers

PaaS orchestrators in PaaS Management Layer are implemented on Java platform with web user interface. Orchestrators are installed on LINUX OS. PI storage is based on Network Attached Storage (NAS). NAS is mounted with read only permissions on Developer and Consumer layers while it is mounted with write permissions on PaaS Management layer. This provides an additional layer of control at storage layer to prevent unauthorized update from a random cloud host. Lab in PaaS Automation Developer Layer consists of 14 servers and is divided into pools to simulate production environment and isolate developers' environments. A total of seven PaaS automation developers worked on the lab environment during the experiment. The developer models and defines automation processes to execute artifacts on the target managed cloud servers. Deployment of database, application server and web application server was simulated. Plans and artifacts were developed to patch and update middleware, database systems and application run time environment in PaaS.

PaaS Consumer Layer consists of 26 tenants hosted in 219 servers provided by IaaS layer. The hosts are integrated with PaaS PO through Host Agents (HA).

HAs are required by PaaS Automation Orchestrator in this MS. Secure HyperText Transfer Protocol (HTTPS) is used for the communication between PaaS Orchestrator and PaaS cloud resources. HAs are started automatically during the booting process of the servers and run with IaaS privileged accounts in the managed consumer servers. The agents digitally trust the orchestrator. Automation artifacts are whitelisted through definitions in the HAs. By that, HAs provide additional layer of control on cloud hosts.

### B. IaaS Management and Security Audit Layers

In IaaS Management Layer, 90% of servers are virtual and managed by hypervisor. The remaining 10% includes bare metals. The bare metals host web applications and multi-tenant platform containers in the experiment. During the provisioning process of an IaaS server, HA is installed automatically as a sub-provisioning plan and PI NAS is mounted automatically. Artifacts definitions in Production Engine are reflected in the HA to achieve automated and direct integration with PaaS orchestrator.

In Security Audit Layer, security policies for DPCs were created. The policies are assigned to PaaS developers and PaaS administrators based on roles. DPCs include PI management commands to add new components or definitions, commands to deploy from Development Engine to Production Engine, and commands to rollout automation definitions from Engines to Cloud HAs. Contents of automation artifacts and definitions are written to Security Events database during the release, deploy or rollout of these automation units. Management of HAs is also achieved securely using

DPCs by PaaS administrator. Events at HAs are monitored and logged in Security Events subsystem.

## VI. ANALYSIS AND DISCUSSION

In this section, critics collected as part of Delphi process are analyzed based on experimental observations. The experiment was conducted in CSP organization. Organizational factors may impose threats to the internal validity of the experiment such as management support and sponsorship.

Critic number one is related to the perceived service quality. PaaS automation cannot be guaranteed by 100% as observed in experiment. QA should minimize the probability of automation failure. Automation developer support is required to fix automation failure based on SLAs.

With respect to critic number two, the following was observed. PaaS automation requires deep understanding of how a platform is deployed and managed. PaaS automation developer needs to consider possible failure scenarios. These scenarios are simulated in lab. PaaS automation developer needs to be an expert in the platform being automated. After the deployment of PaaS automation, the developer is needed to troubleshoot, optimize and fix issues. Upon the release of a new platform version or patch, developer needs to review and update automation. In the experiment, modularity and reusability approaches were adopted. Based on observation, labor technical skills about application platforms were enriched. There was a shift in the type of work being done by the labor. Instead of doing a deployment task manually and sequentially, the work is shifted toward platform deployment automation analysis, PaaS automation development, PaaS automation QA, and PaaS automation lifecycle management. Throughput of one technical labor is increased.

Critic number 3 is related to the perceived relative advantage. Database platform deployment automation was analyzed from human hours' perspectives. To deploy 100 database systems manually, labor needs to work sequentially. Deployment of one database system manually requires 4 hours of work in average. In total, 400 human hours are required to deploy 100 servers. On the other hand, development and QA of automating database system deployment requires an average of 80 hours based on the study experiment. The created value is extended with the deployment of more servers while the initial investment cost from the CSP is the same. Breakeven is achieved after the deployment of the 20th database system. Additional benefit beside human hours saving include service agility. Deployment of the server automatically was done in less than 30 minutes. On the other hand, cost of operating PaaS MS should be considered. In addition, customer support service is required in case of automation failure. Accordingly, financial feasibility study is required to measure return on PaaS automation investment.

Critic number 4 is related to PaaS compatibility with the organization. Policies and procedures were updated to achieve the experiment in the organization. An example for that is handling the use of root user which is a privileged account at LINUX OS system level. By focusing on the com-

mon goal of providing competitive and secure services to customers and with organization management support, legacy policies were updated to fit with the proposed PaaS MS.

Critic number 5 can be related to organization size. The experiment was conducted in organization where OSs, database systems and application servers are supported by different functional units. Each team has a set of privileged accounts to manage its services. In the experiment, it was found that deployment of database requires OS and database privileged accounts. The proposed design described above resolved SoD issues associated with the use of privileged accounts owned by different functional group.

Critic number 6 involves the perceived security risk of PaaS. The unlimited power of PaaS Orchestrator was controlled using the proposed MS. Only authorized artifacts are allowed to be executed by the Orchestrator. Authorization is achieved through the definitions in the HAs. The distribution of definitions from PI to HAs is achieved securely by the MS. Integrity of the definitions and the associated artifacts is monitored through Dual Integrity check subsystem.

With respect to critic number seven, it is not believed that the proposed solution in this paper is the only possible management solution for PaaS. However, different solutions can be proposed and evaluated. Comparison between systems can be discussed in future research papers.

With respect to the last critic, it was confirmed in the experiment that there is an overhead of maintaining application platforms delivered by PaaS. In the experiment, flexibility, reusability, complexity and dependency automation properties [5] were incorporated in automation development to minimize the maintenance overhead. Modular and reusable components represent the logical units of PaaS automation plans. Parameter can be passed to these reusable units to control the use case. Also plan based automation and image based approaches were utilized. By that, a base image of the platform is maintained. Then, new functions, updates and patches are added to the platform using plan artifacts to roll-forward the platform to the desired version.

## VII. CONCLUSION

Based on this study, managing mission critical systems in private PaaS Cloud is practical. The study revealed PaaS adoption factors which can be categorized into organizational, relative advantage, compatibility, security and quality factors. Organizational, compatibility and relative advantage factors are analyzed in the previous section. Security and quality technological factors are incorporated in the MS model and experimented. The proposed model is considered a novel MS for CSPs aiming to provide high level of security and quality standards. The study also gives PaaS cloud users an understanding of how systems can be managed internally by CSPs which should lead to better SLAs in the future.

Comparison with other models and technologies can be done in future research. It is also suggested to validate the model under public cloud and examine the forensic aspects.

## REFERENCES

- [1] M. E. Alex and R. Kishore, "Forensics framework for cloud computing," *Computers & Electrical Engineering*, vol. 60, 2017, pp. 193-205. doi:10.1016/j.compeleceng.2017.02.006
- [2] S. Costache, D. Dib, N. Parlavantzas and C. Morin, "Resource management in cloud platform as a service systems: Analysis and opportunities," *Journal of Systems and Software*, vol. 132, 2017, pp. 98-118. doi:10.1016/j.jss.2017.05.035
- [3] S. N. Deshmukh and H. P. Khandagale, "A system for application deployment automation on cloud environment," *2017 Innovations in Power and Advanced Computing Technologies (i-PACT)*, Vellore, 2017, pp. 1-4. doi:10.1109/IPACT.2017.8245025
- [4] M. Boschetti, V. Baglio, P. Ruiu and O. Terzo, "A Cloud Automation Platform for Flexibility in Applications and Resources Provisioning," *2015 Ninth International Conference on Complex, Intelligent, and Software Intensive Systems*, Blumenau, 2015, pp. 204-208. doi:10.1109/CISIS.2015.29
- [5] J. Wettinger, V. Andrikopoulos, F. Leymann and S. Strauch, "Middleware-Oriented Deployment Automation for Cloud Applications," *IEEE Transactions on Cloud Computing*, vol. 6, no. 4, pp. 1054-1066, 1 Oct.-Dec. 2018. doi:10.1109/TCC.2016.2535325
- [6] J. O. Benson, J. J. Prevost and P. Rad, "Survey of automated software deployment for computational and engineering research," *2016 Annual IEEE Systems Conference (SysCon)*, Orlando, FL, 2016, pp. 1-6. doi:10.1109/SYSCON.2016.7490666
- [7] A. J. Ferrer, D. G. Pérez and R. S. González, "Multi-cloud Platform-as-a-service Model, Functionalities and Approaches," *Procedia Computer Science*, vol. 97, 2016, pp. 63-72. doi:10.1016/j.procs.2016.08.281
- [8] J. Wettinger, V. Andrikopoulos, S. Strauch and F. Leymann, "Characterizing and Evaluating Different Deployment Approaches for Cloud Applications," *2014 IEEE International Conference on Cloud Engineering*, Boston, MA, 2014, pp. 205-214. doi:10.1109/IC2E.2014.32
- [9] J. Wettinger, U. Breitenbücher, O. Kopp and F. Leymann, "Streamlining DevOps automation for Cloud applications using TOSCA as standardized metamodel," *Future Generation Computer Systems*, vol. 56, 2016, pp. 317-332. doi:10.1016/j.future.2015.07.017
- [10] X. Lan, Y. Liu, X. Chen, Y. Huang, B. Lin and W. Guo, "A Model-Based Autonomous Engine for Application Runtime Environment Configuration and Deployment in PaaS Cloud," *2014 IEEE 6th International Conference on Cloud Computing Technology and Science*, Singapore, 2014, pp. 823-828. doi:10.1109/CloudCom.2014.80
- [11] K. Kritikos, T. Kirkham, B. Kryza and P. Massonet, "Towards a Security-Enhanced PaaS Platform for Multi-Cloud Applications," *Future Generation Computer Systems*, vol. 78, Part 1, 2018, pp. 155-175. doi:10.1016/j.future.2016.10.008
- [12] J. O. de Carvalho, F. Trinta, D. Vieira and O. A. C. Cortes, "Evolutionary solutions for resources management in multiple clouds: State-of-the-art and future directions," *Future Generation Computer Systems*, vol. 88, 2018, pp. 284-296. doi:10.1016/j.future.2018.05.087
- [13] Y. Jinzhou, H. Jin, Z. Kai and W. Zhijun, "Discussion on private cloud PaaS construction of large scale enterprise," *2016 IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*, Chengdu, 2016, pp. 273-278. doi:10.1109/ICCCBDA.2016.7529570
- [14] P. Mell and T. Grance, "The NIST Definition of Cloud Computing," National Institute of Standards Technology, Gaithersburg, MD, USA, Technical Report, SP 800-145, 2011. doi:10.6028/NIST.SP.800-145
- [15] M. Kostoska, I. Chorbev and M. Gusev, "Creating portable TOSCA archive for iKnow University Management System," *2014 Federated Conference on Computer Science and Information Systems*, Warsaw, 2014, pp. 761-768. doi:10.15439/2014F311
- [16] L. Roderio-Merino, L. M. Vaquero, E. Caron, A. Muresan and F. Desprez, "Building Safe PaaS Clouds: A Survey on Security in Multitenant Software Platforms," *Computers & Security*, vol. 31, Issue 1, 2012, pp. 96-108. doi:10.1016/j.cose.2011.10.006
- [17] R. Dukaric and M. Juric, "BPMN extensions for automating cloud environments using a two-layer orchestration approach," *Journal of Visual Languages and Computing*, vol. 47, pp. 31-43. doi:10.1016/j.jvlc.2018.06.002



# Project Management Tasks in Agile Projects: A Quantitative Study

Gloria J. Miller

Managing Consultant  
maxmetrics

Heidelberg, Germany

<https://orcid.org/0000-0003-2603-0980>

**Abstract**—Recent studies have confirmed the efficacy of agile methodologies in project success. However, can projects skip several project management tasks and still deliver the expected results? How are traditional project managers engaged in agile projects? The results from this study quantify subjective and theoretical speculation on who performs the project management tasks in agile projects. Project managers are engaged in agile projects and the team, the product owner, and project sponsor are significantly involved in project management tasks. The agile coach is not a substitute for the project manager. The study identifies that agile and traditional methodologies should be updated to clarify team, product owner, and agile coach responsibilities.

## I. INTRODUCTION

WHILE the adoption of agile project management methodologies is widespread [1], the project management tasks in agile projects are uncertain and the lack of clarity is causing confusion in practice [2, 3, 4]. Agile methodologies provide events, processes, and artifacts that should allow projects to be flexible to change and deliver results in an iterative, incremental fashion. Some of the most popular agile project management frameworks, such as scrum, extreme programming (XP), and lean/kanban processes do not explicitly include a project manager role or project management tasks. For example, scrum includes three roles: a scrum master, product owner, and the team.

Hobbs and Petit [4] identified 827 articles that compare or integrate agile project management with project management practices. However, not once did the research reference the effect of agile methodologies on the project manager role. Other agile research recognizes the confusion and conflict caused by the lack of recognition of the project manager in agile methodologies [5]. There is some speculation that the project manager is better suited to take over the product owner role [2] than the agile coach. While other research starts with the premise that there is an agile project manager who is a facilitator or coach [6, 7, 8, 9]. Even the *Agile Practice Guide* issued by the Project Management Institute (PMI) in 2017 states that the “role of the Project Manager in an agile project is somewhat unknown” [10, p. 37]. In agile methodologies, some but not

all of the typical project management responsibilities have been distributed to other roles [9].

The success rate for agile methodologies is on par with, if not better than, those managed under a traditional methodology [1]. Thus, if agile methodologies are followed rigorously and exclude a project manager, then maybe the project manager role and some project management tasks are obsolete. Shastri, Hoda and Amor [11] found that the project manager role does still exist in agile projects. However, the study left open the questions for why and what activities the project manager performs. “The implementation of agile methods can have a very significant impact on the role of the project manager, but a better understanding of the circumstances under which the project manager role changes and how it changes is needed” [4, p. 11]. Other studies have investigated agility in projects or the effects of specific practices on the success of projects applying agile methodologies [1, 12, 13, 14]. Nevertheless, the topic of who performs what project management activities in agile projects remain unanswered. Specifically, *how are traditional project managers engaged in agile projects? Who executes what project management tasks in projects applying agile methodologies?*

First, we review the literature for understanding agile methodologies and the roles and tasks involved in the project work using agile and plan-driven methodologies. We map the responsibilities from an agile methodology to the project management knowledge areas and processes from the International Organization for Standardization (ISO) project management standard. Then, we define and conduct a survey to understand what roles perform which project management tasks. Finally, we quantitatively analyze the survey results and answer the research questions.

The results from this study quantify subjective and theoretical speculation on who performs the project management tasks in agile projects. The results contribute to the project management knowledge on agile methodologies by widening the perspective on the role of people, specifically the project manager, in agile projects. The next section reviews the literature and describes the research methodology, results, and discussions. The final sections of the study discuss the conclusions and implications.

## II. LITERATURE REVIEW

Traditional, waterfall or plan-driven methodologies follow a stage-gate or phased lifecycle. These methods have in common the creation of an upfront plan, where the time is limited with the limitation and termination conditions known from the beginning [15]. The methodologies and frameworks for traditional projects are codified in the project management standards and frameworks, such as “ISO 21500:2012, Guidance on Project Management” [16], *APM Body of Knowledge 6th Edition* [17], and *A Guide to the Project Management Body of Knowledge (PMBOK guide)* [18].

The *Agile Manifesto* is a set of four values and 13 principles that provide a framework for managing technology projects in a flexible way that responds to dynamic project situations [1, 19, 20]. There are at least 13 methodologies or frameworks that can be considered to follow the values and principles described in the *Agile Manifesto*. Each agile methodology has its own set of roles, rules, events, and practices; however, in general, they encourage iterative and incremental development life cycles, self-organizing teams, and evolutionary product development. Scrum, XP, lean, and Kanban are the most frequently referenced agile frameworks in surveys on agile adoption and in the project management literature [11].

The project manager is the authorized person who leads and manages project activities and is accountable for project completion [21]. The role is defined in ISO, APM, and PMI project management standards and frameworks. In addition, the standards describe knowledge areas that are expected of a project manager and processes that should be led or executed as part of managing a project. The project management literature agrees that the project manager has the sole responsibility for planning and managing projects [21]. The project manager should direct the performance of the planned project activities and manage the various technical, administrative and organizational interfaces within the project [16, 18]. The project manager role is not explicit in the agile methodology. Noll, Razzak, Bass and Beecham [2] found that the scrum master, a coaching role in the scrum agile methodology, combines project management activities in practice. However, there was tension created since the scrum master should balance management activities with coaching the team. The inherent suggestion in studies about project managers in agile projects is that the leadership style or skills, knowledge, personal attributes and behavior of the project manager must be adapted [8, 9, 22, 23].

There are 39 processes in 10 subject areas that cover five process groups described for the project management role in the ISO project management standard [16, 20]. The principles and values from the *Agile Manifesto* offer a framework on how people should work [20]. Consequently, the manifesto does not explicitly establish who should do the work. Binder, Aillaud and Schilli [20] correlated the 12 agile principles to the ISO processes to establish a hybrid model for managing agile projects. They identified gaps and

practice modifications that would have to occur to effectively manage agile projects.

Project management success evaluates performance against the time, budget, and quality constraints of the project, also known in modern terminology as project efficiency [1]. Agile projects have been shown to have similar performance as traditional methodologies for project efficiency [1]. Nevertheless, the only success measure given for agile projects is working software [19].

The project management tasks in plan-driven methodologies are centralized to the project manager role. For agile methodologies, some of the project management responsibilities are inherent in the methods, while the project management activities or tasks are not explicitly identified. Studies on project managers in agile projects have identified conflicts with other agile roles or assume project manager must adapt to manage agile projects.

## III. METHODOLOGY

The research used a literature review to define the project boundary and project management authority, tasks, and roles in projects using plan-driven or agile methodologies. Peer-reviewed journal publications from 2006 to 2018 were evaluated to identify project managers activities in agile projects. A web-based survey was used to collect data on the roles engaged in projects and the project management tasks they perform. A quantitative analysis method was used to explore the difference between the theoretical and practical applications of project management tasks.

The survey sample comprised 120 usable responses as follows: 33% of the respondents had a project manager role; 11% were program managers; 9% were from a project management office; 9% agile coaches or scrum masters; 8% product owners; 24% project team members from IT, business, software vendors, or others not in the selection list; 3% project sponsors, and less than 2% others. The organizations sponsoring the projects were spread throughout 20 different industries. The participants were balanced across geographic regions: Europe (25%), Asia (19%), Latin America, and the Caribbean (18%), North America (16%), and Oceania (3%). Most of the projects started in the last five years (81%), lasted more than one year (56%), and had less than 21 team members (81%).

The SAS Studio (Release 3.6, basic edition) was used to perform the statistical analysis and produce the tables and figures. The responses were checked for scope, completeness, consistency, ambiguity, missing data, extreme responses, outliers, and leverage. No bias was found, and the data were reliable and valid. The number responses is comparable with other agile studies: [9] with 20 agile practitioners, and [11] with 97 and [6] with 32 survey respondents. Thus, the data were considered valid for further analysis. The descriptive statistics, mean ranking, Wilcoxon score, and correlation analysis were used to explore the characteristics, establish the validity and reliability, and explain the relationship between the variables. The descriptive statistics provide insight into the content and

structure of the projects, the involvement of the different project roles, and the relationship between the involved roles and the methodology. The hypothesis testing uses regression analysis to explain the relationships between the roles that performed project management tasks and project efficiency.

#### IV. DATA ANALYSIS AND RESULTS

The analysis was performed for the four topic areas of project efficiency, methodology, roles, and tasks. Table I includes the mean comparison between methodology types; significant differences are based on the Kruskal Wallis being less than .05 for 95% confidence. Project budget, time, and quality performance were combined into a single project efficiency variable by taking the mean value of the variables. There is no significant difference between the composite *project efficiency* measure or the individual performance measures by methodology type. Different than some other research, the mixed and plan-driven methods have a higher mean than the agile methods [1].

Scrum and waterfall are the top methodologies at 22% and 20%, respectively. From the agile methodologies, scrum combined with the scrum/XP is the most widely used at 24%. This is consistent with other studies that found scrum to be the most popular agile methodology and agile methodologies are in wide-spread use [1, 11]. The methodology types and methodologies were not significantly correlated with any of the individual performance measures or the project efficiency factor.

The project manager role was involved in 67% of the projects, including 58% of the agile projects, 82% of the mixed methodology projects, and 79% of the plan-driven projects. The agile coach role was included in 35% of all projects and the product owner role in 42% of all projects. There was no significant difference between methodologies for the other roles. The agile coach, product owner, and team combination – a full scrum team – was not present in all scrum-related projects. This implies that the scrum methodology is not being rigorously applied in practice. The project manager was more prominent in the plan-driven and mixed methodology, and the agile coach and product owner were more prominent in the agile methodologies. Otherwise, there was no significant difference in the roles between methodologies.

Project management tasks are performed in all types of methodologies with no significant difference. Overwhelmingly, the project manager is responsible for the project management tasks in all types of methodologies. This involvement is significant for managing the team and stakeholders, identifying risks, establishing the team, and controlling resources. On the other hand, the team is more often identified as being involved in assessing, treating, and controlling risks in plan-driven methodologies, while it is more often the product owner in agile methodologies. The product owner is strongly represented in managing stakeholders. The team and agile coach are not significantly engaged in this task. The project team is involved in

TABLE I  
MEAN AND KRUSKAL-WALLIS TEST

Theme	Measurement Item	Mean			Kruskal Wallis
		Agile	Plan-driven	Mixed	p-Value
Demo graphics	Team Size	3.57	3.06	3.41	0.26
	Duration	2.42	2.29	2.48	0.82
Performance	Requirements	5.01	5.06	5.31	0.60
	Project Eff	4.43	4.39	4.89	0.34
	Budget	4.33	4.35	4.97	0.48
	Time	3.95	3.76	4.38	0.43
	Overall	3.59	3.53	3.66	0.78
Roles	Team Role	0.74	0.76	0.79	0.87
	Project Mgr.	0.58	0.82	0.79	0.04
	Product Own	0.53	0.41	0.14	0.00
	Sponsor	0.38	0.53	0.38	0.50
	Agile Coach	0.49	0.18	0.10	0.00
Tasks	Risk Mgt	3.22	3.41	3.22	0.88
	Procurement Mgt	2.65	2.5	3.17	0.34
	Resource Mgt	2.65	2.55	2.30	0.52

procurement in plan-driven methodologies and not at all involved in agile methodologies.

Fig. 1 combines the qualitative results from the literature review with the quantitative results and provides a consolidated view of project management responsibilities for agile projects. The rows represented the 39 ISO processes grouped into the 10 ISO subject areas, the columns represent the method or the project roles considered in the study, and the color represents the relative degree the processes are executed. For example, the integration subject area includes 6 processes: 2 map to 2 agile principles, 4 to 1 principle, and 1 does not map to any principles. The agile coach role maps to 3 processes for the subject area, the product owner to 2, the team to 3, the project manager to 6, and the sponsor to 1.

#### VI. DISCUSSION

First, while agile project methodologies are popular; nevertheless, traditional methodologies continue to be in widespread use. No matter the methodology, the project management tasks as identified in the ISO standard for managing projects remain relevant. Project managers are engaged in agile projects to a larger degree than agile coaches or product owners. Project managers continue to perform management tasks and not only act as a “gatekeeper” as described by Taylor [3]. In this study, the sponsor and product owner undertook some management activities, while the agile coach did not. Thus, this partially supports the proposition by Noll, Razzak, Bass and Beecham [2] that assigning a former project manager to the product owner role rather than the scrum master role will result in a higher degree of project success. In practice, the project manager focuses on team management and risk identification

tasks, while the product owners focus on the scope and stakeholder management. The product owner is responsible for the scope before the project start, during the project, and after the project is completed. A project manager is a transitory role and is not typically engaged in the market and withdrawal phases of a product lifecycle. Stated differently, the boundary for the product owner is the product features, while for the project manager, it is the project processes. Thus, the pairing of sponsor and product owner is a logical for the long-term product success and the sponsor and project manager for short-term project success.

From the study, it was clear that the agile coach has a much more limited set of tasks than the project manager. A sizable percentage of projects succeeded without an agile coach. The agile coach has two primary responsibility areas: developing the team and supporting all stakeholders to understand and apply the methods. In this regard, the management style suggested for an agile coach is that of a good leader, facilitator, or coach [22]. This corresponds with studies that argue project managers that can practice a facilitator leadership style could lead agile projects [8, 9, 22, 23]. However, the finding by Noll, Razzak, Bass and Beecham [2] that there is likely to be conflict within the project related to the delegation and management styles required by the different sets of responsibilities is a valid reason to separate the project manager and coaching roles. Furthermore, the results show that the agile coach was relevant to project efficiency but not to any of the project management tasks under investigation.

## VII. CONCLUSIONS

The tasks that would typically be carried out by a project manager continue to be practiced while they are not (explicitly) addressed by the agile methodologies. The team, product owner, and project sponsors are taking on the informal role on some project management tasks; the project manager continues to be engaged, albeit with an altered task distribution and leadership style.

The practical implication is that project sponsor should consider the project manager an essential role for all project types and not assign the project management activities to the sponsor or the product owner. The agile methodology authors should update their practices to identify the role the project team takes in assessing, treating, and controlling risks and in planning and managing suppliers. Furthermore, they should reflect the project management tasks that may be outside of the team operations but necessary for project sponsors or project managers to execute. Fig. 1 provides a guideline for mapping specific project roles to project management activities. Furthermore, since the rigorous application of the method is responsible for some typical project management activities, care should be taken to consider project governance when tailoring agile methods.

The agile coach is a key role that can improve the productivity of project operations. Thus, the role should be formalized into traditional methodologies as a role separate

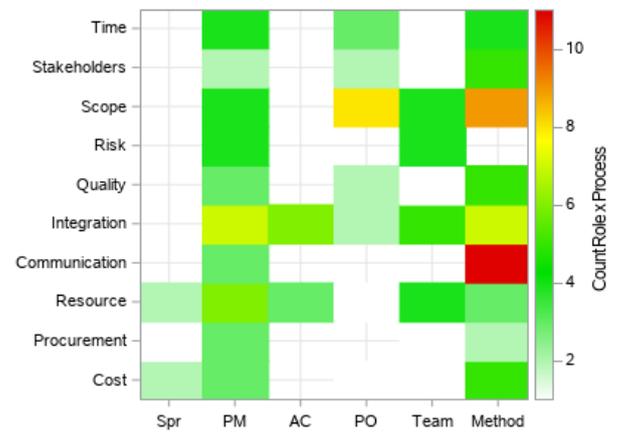


Fig 1. Heatmap of ISO process groups by scrum roles and agile methods based upon relative degree tasks, covered by the role after the analysis.

Legend: Spr-Sponsor; PM-Project Manager; AC-Agile Coach; PO-Product Owner; Method-Agile Principles

from the project manager. Finally, researchers on agile methodologies should consider the participants within the projects in future studies. We determined that less than half the projects following a scrum methodology consist of all the scrum roles. Thus, the actual results from project studies may inappropriately contribute the successful outcomes to the methodology.

The results from this study quantify subjective and theoretical speculation on who performs the project management tasks in agile projects. The results of this study contribute to the project management knowledge on agile methodologies by widening the perspective on the role that people, specifically the project manager. Although agile methodologies are in widespread use, this information was missing in practice and under-researched in academia. Conversely, the research is limited in several ways. The results of this study are not generalizable beyond the methodologies studied in this research; specifically, software development projects have been the most active in applying agile methodologies. There were no measures to determine whether the project type skewed or biased the results. The research did not have financial or factual data to measure performance. Thus, it could only evaluate the perception of project performance as judged by the participants. Furthermore, the findings are limited due to the small sample size. Future research could focus on a qualitative study of the agile project organizations or seek to quantify the engagement of the separate roles.

## REFERENCES

- [1] P. Serrador and J. K. Pinto, "Does Agile work? — A quantitative analysis of agile project success," *International Journal of Project Management*, vol. 33, no. 5, pp. 1040-1051, July 2015, <https://doi.org/10.1016/j.jiproman.2015.01.006>.

- [2] J. Noll, M. A. Razzak, J. M. Bass, and S. Beecham, "A Study of the Scrum Master's Role," in Felderer M., Méndez Fernández D., Turhan B., Kalinowski M., Sarro F., Winkler D. (eds). *Product-Focused Software Process Improvement. PROFES 2017*, Springer, Cham, 2017.
- [3] K. J. Taylor, "Adopting Agile software development: the project manager experience," *Information Technology & People*, vol. 29, no. 4, pp. 670-687, 2016, <https://doi.org/10.1108/ITP-02-2014-0031>.
- [4] B. Hobbs and Y. Petit, "Agile Methods on Large Projects in Large Organizations," *Project Management Journal*, vol. 48, no. 3, pp. 3-19, 2017, <https://doi.org/10.1177/875697281704800301>.
- [5] K. Dikert, M. Paasivaara, and C. Lassenius, "Challenges and success factors for large-scale agile transformations: A systematic literature review," *Journal of Systems and Software*, vol. 119, no. pp. 87-108, Sept 2016, <https://doi.org/10.1016/j.jss.2016.06.013>.
- [6] Z. Mansor, N. H. Arshad, S. Yahya, and R. Razali, "The competency of project managers in managing agile cost management," *Advanced Science Letters*, vol. 22, no. 8, pp. 1930-1934, 2016, <https://doi.org/10.1166/asl.2016.7750>.
- [7] K. Conboy and L. Morgan, "Combining open innovation and agile approaches: implications for IS project managers," 2010.
- [8] K. Sutling, Z. Mansor, S. Widyarto, S. Letchmunan, and N. H. Arshad, "Agile project manager behavior: The taxonomy," in 2014 8th. Malaysian Software Engineering Conference (MySEC), 2014, pp. 234-239, <https://doi.org/10.1109/MySec.2014.6986020>.
- [9] K. Sutling, Z. Mansor, S. Widyarto, S. Letchmunan, and N. H. Arshad, "Understanding of project manager competency in agile software development project: The taxonomy," in *Lecture Notes in Electrical Engineering*. vol. 339, ed, 2015, pp. 859-868, [https://doi.org/10.1007/978-3-662-46578-3\\_102](https://doi.org/10.1007/978-3-662-46578-3_102).
- [10] PMI, "Agile Practice Guide," ed: Project Management Institute, Inc., 2017.
- [11] Y. Shastri, R. Hoda, and R. Amor, "Does the 'Project Manager' Still Exist in Agile Software Development Projects?," in 2016 23rd Asia-Pacific Software Engineering Conference (APSEC), 2016, pp. 57-64, <https://doi.org/10.1109/APSEC.2016.019>.
- [12] K. M. Mayfield, "Project managers' experience and description of decision uncertainty associated with the agile software development methodology: A phenomenological study," 3427057 Ph.D., Capella University, Ann Arbor, 2010.
- [13] J. Sheffield and J. Lemétayer, "Factors associated with the software development agility of successful projects," *International Journal of Project Management*, vol. 31, no. 3, pp. 459-472, March 2013, <https://doi.org/10.1016/j.ijproman.2012.09.011>.
- [14] V. Liubchenko, "A review of agile practices for project management," in 2016 XIth International Scientific and Technical Conference Computer Sciences and Information Technologies (CSIT), 2016, pp. 168-170, <https://doi.org/10.1109/STC-CSIT.2016.7589897>.
- [15] R. A. Lundin and A. Söderholm, "A theory of temporary organization," *Scandinavian Journal of Management*, vol. 11, no. pp. 437-455, 1995, [https://doi.org/10.1016/0956-5221\(95\)00036-U](https://doi.org/10.1016/0956-5221(95)00036-U).
- [16] ISO, "ISO 21500: 2012 Guidance on project management," in International Organization for Standardization, ed, 2012.
- [17] APM, "APM body of Knowledge 6th Edition," ed: Association for Project Management, 2012.
- [18] PMI, "A Guide to the Project Management Body of Knowledge (PMBOK Guide)," Sixth Edition ed. Newtown Square, PA: Project Management Institute, Inc., 2017.
- [19] <http://agilemanifesto.org/>, "Manifesto for Agile Software Development," in *Agile Manifesto*, ed, 2001.
- [20] J. Binder, L. I. V. Aillaud, and L. Schilli, "The Project Management Cocktail Model: An Approach for Balancing Agile and ISO 21500," *Procedia - Social and Behavioral Sciences*, vol. 119, no. pp. 182-191, March 2014, <https://doi.org/10.1016/j.sbspro.2014.03.022>.
- [21] O. Zwikael and J. R. Meredith, "Who's who in the project zoo? The ten core project roles," *International Journal of Operations & Production Management*, vol. 38, no. 2, pp. 474-492, 2018, <https://doi.org/10.1108/IJOPM-05-2017-0274>.
- [22] N. A. Bonner, "Predicting Leadership Success In Agile Environments: An Inquiring Systems Approach," *Academy of Information and Management Sciences Journal*, vol. 13, no. 2, pp. 83-103, Sep 2010.
- [23] H. Yang, S. Huff, and D. Strode, "Leadership in software development: Comparing perceptions of agile and traditional project managers," *AMCIS 2009 Proceedings*, p. 184, 2009.



# 25<sup>th</sup> Conference on Knowledge Acquisition and Management

**K**NOWLEDGE management is a large multidisciplinary field having its roots in Management and Artificial Intelligence. Activity of an extended organization should be supported by an organized and optimized flow of knowledge to effectively help all participants in their work.

We have the pleasure to invite you to contribute to and to participate in the conference "Knowledge Acquisition and Management". The predecessor of the KAM conference has been organized for the first time in 1992, as a venue for scientists and practitioners to address different aspects of usage of advanced information technologies in management, with focus on intelligent techniques and knowledge management. In 2003 the conference changed somewhat its focus and was organized for the first under its current name. Furthermore, the KAM conference became an international event, with participants from around the world. In 2012 we've joined to Federated Conference on Computer Science and Systems becoming one of the oldest event.

The aim of this event is to create possibility of presenting and discussing approaches, techniques and tools in the knowledge acquisition and other knowledge management areas with focus on contribution of artificial intelligence for improvement of human-machine intelligence and face the challenges of this century. We expect that the conference&workshop will enable exchange of information and experiences, and delve into current trends of methodological, technological and implementation aspects of knowledge management processes.

## TOPICS

- Knowledge discovery from databases and data warehouses
- Methods and tools for knowledge acquisition
- New emerging technologies for management
- Organizing the knowledge centers and knowledge distribution
- Knowledge creation and validation
- Knowledge dynamics and machine learning
- Distance learning and knowledge sharing
- Knowledge representation models
- Management of enterprise knowledge versus personal knowledge
- Knowledge managers and workers
- Knowledge coaching and diffusion
- Knowledge engineering and software engineering
- Managerial knowledge evolution with focus on managing of best practice and cooperative activities
- Knowledge grid and social networks

- Knowledge management for design, innovation and eco-innovation process
- Business Intelligence environment for supporting knowledge management
- Knowledge management in virtual advisors and training
- Management of the innovation and eco-innovation process
- Human-machine interfaces and knowledge visualization

## EVENT CHAIRS

- **Hauke, Krzysztof**, Wroclaw University of Economics, Poland
- **Nycz, Malgorzata**, Wroclaw University of Economics, Poland
- **Owoc, Mieczyslaw**, Wroclaw University of Economics, Poland
- **Pondel, Maciej**, Wroclaw University of Economics, Poland

## PROGRAM COMMITTEE

- **Abramowicz, Witold**, Poznan University of Economics, Poland
- **Andres, Frederic**, National Institute of Informatics, Tokyo, Japan
- **Bodyanskiy, Yevgeniy**, Kharkiv National University of Radio Electronics, Ukraine
- **Chmielarz, Witold**, Warsaw University, Poland
- **Christozov, Dimitar**, American University in Bulgaria, Bulgaria
- **Jan, Vanthienen**, Katholike Universiteit Leuven, Belgium
- **Mercier-Laurent, Eunika**, University Jean Moulin Lyon3, France
- **Sobińska, Małgorzata**, Wroclaw University of Economics, Poland
- **Surma, Jerzy**, Warsaw School of Economics, Poland and University of Massachusetts Lowell, United States
- **Vasiliev, Julian**, University of Economics in Varna, Bulgaria
- **Zhu, Yungang**, College of Computer Science and Technology, Jilin University, China

## ORGANIZING COMMITTEE

- **Hołowińska, Katarzyna**
- **Przysucha, Łukasz**, Wroclaw University of Economics



# Analysis of Relationship between Personal Factors and Visiting Places using Random Forest Technique

Young Myung Kim

Department of Computer Engineering, Hongik University  
Seoul, Republic of Korea  
Email: dudaud0205@gmail.com

Ha Yoon Song

Department of Computer Engineering, Hongik University  
Seoul, Republic of Korea  
Email: hayoon@hongik.ac.kr

**Abstract**—There has been research regarding relationship between human personalities and visiting places using Big Five Factor (BFF). However, other factors such as Social media usage, Hobby, Gender, Age, and Religion and so on are regarded as also major factors which effects the choice of visiting place of a person. Using questionnaire designed by authors, these factors as well as BFF were prepared for this research. The visiting places were collected by a smartphone app called SWARM and classified in 10 categories. In sum, personal data of 34 participants had been collected for several months. To figure out the relationship between these factors and visiting places, random forest technique of ensemble method was used.

## I. INTRODUCTION

PRIOR researches show that human personality and favorite visiting place have considerable relationship [1] [2] [3] [4]. However, there has been long belief that other than personalities, personal factors effect the selection of visiting location. To prove this belief, we collected personal factors other than personality from survey. Gender, Age, Marital Status, Religion, Salary, Vehicles, usage of SNS, Job, Educational Level, Frequency of travel for a year, Time spent on SNS per day, sort of hobby. Using Big Five Inventory (BFI), we collected person's Big Five Factor (BFF) Total 34 participants provided their personal data and location data. To collect location data, a smartphone application called SWARM is used and the duration of collection was up to six months. The method to analyze these data is Random Forest which is ensemble learning.

### A. Random Forest

Random Forest is suggested by Leo Breiman in 2001 [5]. Random Forest shows good performance and high accuracy in general and without overfitting. It can handle many input features and resistant to noise. In addition, the degree of effect of input feature can be numerically represented as importance value. We considered Random Forest as a suitable method for our research.

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (NRF-2019R1F1A1056123).

### B. Big Five Factors (BFF)

BFF is a factor of personality suggested by P.T. Costa and R.R. McCrae in 1992 [6]. It has five factors of Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. Set of questionnaires is answered by participants and each five factors will be valued as a score from 0 to 5 points. Since BFF can numerically represent conceptual human personality, many of research adopts BFF [7] [8] [9] [10] [11] [12].

### C. SWARM Application

SWARM application is used to collect geo-positioning data installed on smartphones [13]. Users actively check in visited places with SWARM. These actively collected location data are used as part of our analysis.

In section II, we will discuss random forest and our purpose in this research. Section III will show details of the data. The handling of personal factors and location categories will be discussed. Section IV will show results of analysis by Random Forest and evaluate the results. Section V will conclude this research with future works.

## II. RANDOM FOREST TECHNIQUE

### A. Ensemble

Ensemble is a technique which combines various machine learning models to generate powerful model. Random Forest is a sort of ensemble technique and has decision tree as its base model. Especially, Random Forest and gradient boosting have proven as useful method for classification and regression of various data set. These two distinguished models have base element of decision tree.

### B. Decision Tree

Decision tree is a widely used model for classification and regression. Basically, decision tree is a consequence of yes-no question of leaning process toward the final decision. For example, to classify bear, pigeon, penguin, dolphin with the smallest number of questions, several sequences of question are introduced. The first question to classify two animals is: "Does it have wings?" Then the second question is: "can

it fly?" Then pigeon and penguin can be classified. In case there is no wing, the following question will be: "Does it have fin", and dolphin and bear can be separated. These questions are called as test in machine learning. And decision tree is consisted as nodes for test and edge connected to the following test. In case of machine learning, continuous values can be used instead of yes-no question. In this case, test can be in a form that is feature  $i$  bigger than value  $a$ .

### C. Bootstrap Aggregating (Bagging)

Random Forest creates bootstrap samples of data to create several independent decision trees. Bootstrap samples are random choices of data by allowing redundancy. The size of the dataset is the same as the original dataset. Some data will be missing from the bootstrap sample and some data may be duplicated [14].

The disadvantage of the decision tree is that it can be overfitted to the training data whereas Random Forest can handle this problem. Random Forest is a bundle of different decision trees. Each decision tree is relatively good at prediction but can be overfitted in the training data. However, if we create many of decision trees and average its results, the prediction performance of the decision tree can be enhanced by reducing the overfitting. In addition, each branch of the decision tree uses a subset of different features because only part of the features is used in each node. This method makes all the decision trees in the Random Forest different from each other. Random Forest predicts with results from each decision tree. For the regressions used in this study, average of each result is used to make the final prediction.

Random Forest is one of widely used machine learning algorithm with excellent performance. It is strong in noise, works well even without much hyperparameter tuning, and does not need to scale data. It also works well on very large datasets and can parallelize the train simply. It is also appropriate to deal with many input features [15]. We can also know the value importance of the input value that affects the result. Due to this advantage and performance, a Random Forest was used for this study.

## III. PERSONAL FACTORS AND LOCATION CATEGORIES

### A. Personal Factors

Many of research adopts BFF as a measure of personality suggested by McCrae and Costa. [6] The five factors are Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. Each factor are measured as numerical numbers so that factors can be easily applied to training process. Table I shows BFF of several participants. We can figure out personality of a person through these values. Person with high Openness is creative, emotional and interested in arts. Person with high Conscientiousness is responsible, achieving, and restraint. Person with high Agreeableness is agreeable to other person, altruistic, thoughtfulness and modesty. While person with high Neuroticism is sensitive to stress, impulsive, hostile and depressed. For example, as shown in table I, person 4 is creative, emotional, responsible, restraint. Also considering

TABLE I: BFF of Participants

	O	C	E	A	N
Person 1	3.3	3.9	3.3	3.7	2.6
Person 2	2.7	3.2	3.2	2.7	2.8
Person 3	4.3	3.1	2.3	3.2	2.9
Person 4	4.2	4.3	3.5	3.6	2.6
Person 5	4	3.7	4	3.9	2.8
Person 6	3.8	4	3.1	3.8	2.3
Person 7	3.2	3.2	3.5	3.3	3.5
Person 8	2.8	3.8	3.8	3.3	2.3
Person 9	3.4	3.6	3.5	3.6	3.1
Person 10	3	3.6	2.5	3	3
Person 11	4.1	3.8	3.8	2.8	3
Person 12	3.1	3	2.8	3	2.8
Person 13	3.3	3.2	3.5	2.6	2.6
Person 14	3.7	3.3	3.6	3.8	3.5
Person 15	2.4	3.7	3	2.8	2.6
Person 16	3.4	3.2	3.0	3	2.6
Person 17	3.9	3.3	3.5	2.9	2.8

TABLE II: Personal Factors: Person 1

Personal Factors	Value
Age	2
Job	1
Marriage	2
The highest level of education	2
Major	4
Religion	1
Salary	2
Vehicles	4
Commute time	3
the frequency of a year's journey	2
SNS usage status	1
Time spent on SNS per day	3
cultural life	3
Openness	3.3
Conscientiousness	3.9
Extraversion	3.3
Agreeableness	3.7
Neuroticism	2.6

person 4's Neuroticism, person 4 is not impulsive and resistant to stress. The personality shown in table I will be used our experimental basis with other personal factors.

In the table II, the number corresponding to the response is as follows:

#### Age

1: 10s, 2: 20s, 3: 30s, 4: over 40s

#### Job

1: students, 2: administrative position, 3: expert, 4: an engineer, 5: office job, 6: service, sales position, 7: a functional worker, 8: equipment maneuvering and assembly engineer, 9: simple laborer

cf. Occupational classifications include the International Standard Classification of Occupation (ISCO) [16].

**Marriage**

1: married, 2: single

**The highest level of education**

1: middle school graduate, 2: high school graduate, 3: college graduate, 4: master, 5: doctor

**Major**

1: humanities, 2: sociology, 3: pedagogy, 4: engineering, 5: nature, 6: medicine and pharmacology, 7: art, music and physical education

**Religion**

1: no religion, 2: Christianity, 3: Catholic, 4: Buddhism

**Salary**

1: Less than 500 USD, 2: 500 USD to 1,000 USD, 3: 1,000 USD to 2,000 USD, 4: 2,000 USD to 3,000 USD, 5: over 3,000 USD

**Vehicles**

1: walking, 2: bicycle, 3: car, 4: public transport

**Commute time**

1: less than 30mins, 2: 30mins to 1h, 3: 1h to 2h, 4: over 2 hours

**The frequency of a year's journey**

1: less than one time, 2: 2 to 3 times, 3: 4 to 5 times, 4: over six times

**SNS usage status**

1: Use, 2: Not use

**Time spent on SNS per day**

1: less than 30 mins, 2: 30 mins to 1 hour, 3: 1 hour to 3 hours, 4: over 3 hours

**Cultural life**

1: static activity, 2: dynamic activity, 3: both

In case of Person 1, a number of personal factors are coming from 20s, such as students, single, a high school graduate, engineering, no religion, income in 500USD to 1000USD, public transport, commute in 1 to 2hours, two or three travels per year, one to three hours spent for social media per day, and both dynamic and static cultural life.

**B. Location Category Data**

Location Category data was used as Label (target data) for the supervised learning, Random Forest. The location category data is checked in to the visiting places using the SWARM application. Afterwards, the number of visits and visiting places were identified from web page of SWARM. Part of the location data of person 16 is shown in the table III.

TABLE III: Sample Location Data: Person 16

location	Count of Visit
Hongik Univ. Wowkwan	19
Hongik Univ. IT Center	7
Kanemaya noodle Restaurant	3
Starbucks	3
Hongik Univ. Central Library	8
Coffesmith	2
Daiso	3

The data collected were classified into 10 categories. Table IV shows the classification of person 16's location data into a category.

TABLE IV: Sample Classification of Locations: Person 16

Category	location	Visiting Ratio
Foreign Institutions	0	0
Retail Business	6	0.04
Service industry	6	0.04
Restaurant	29	0.1933
Pub	2	0.0133
Beverage Store	26	0.1733
Theater and Concert Hall	4	0.0267
Institutions of Education	62	0.4133
Hospital	6	0.04
Museum, Gallery, a historical site, tourist spots	9	0.06

To input categorized location data to Random Forest, visiting ratio of location categories are used as labels. The formula is as follows.

$$Visiting\_Ratio = \frac{\text{count\_of\_visit\_to\_location}}{\text{total\_count\_of\_visits}}$$

**IV. ANALYSIS OF RESULTS**

By analyzing data using random forest, you can see value importance, which is the degree of how each feature affects the prediction. Table V shows summary of result for each Location Category such as Symmetric Mean Absolute Percentage Error (SMAPE), Accuracy and the top five feature's importance values with the most impact. Table V abbreviated location category.

**FI:** Foreign Institutions

**RB:** Retail Business

**SI:** Service Industry

**BS:** Beverage Store

**TC:** Theater and Concert Hall

**IE:** Institutions of Education

**MG:** Museum, Gallery, historical sites and tourist spots

The result of the experiment randomly selected one of the decision trees is present in Fig. 1. The unbiased and well-made decision tree is found as shown in Fig. 2 when it has label of Restaurant. Several significant value importance graphs with meaningful accuracy are also shown Fig. 3.

**A. Discussion about Low Prediction Accuracy**

First, we analyzed the reason of very low accuracy, especially for foreign institutions and hospital. This is just because of shortage of data, since most of participants rarely went to foreign instruments. A handful of people have visited international airport only once or twice while traveling abroad.

It would have been difficult to predict because the person who went to foreign institutions lacked data. Hospital shows similar situation. Hospital is not a place to go by a person's

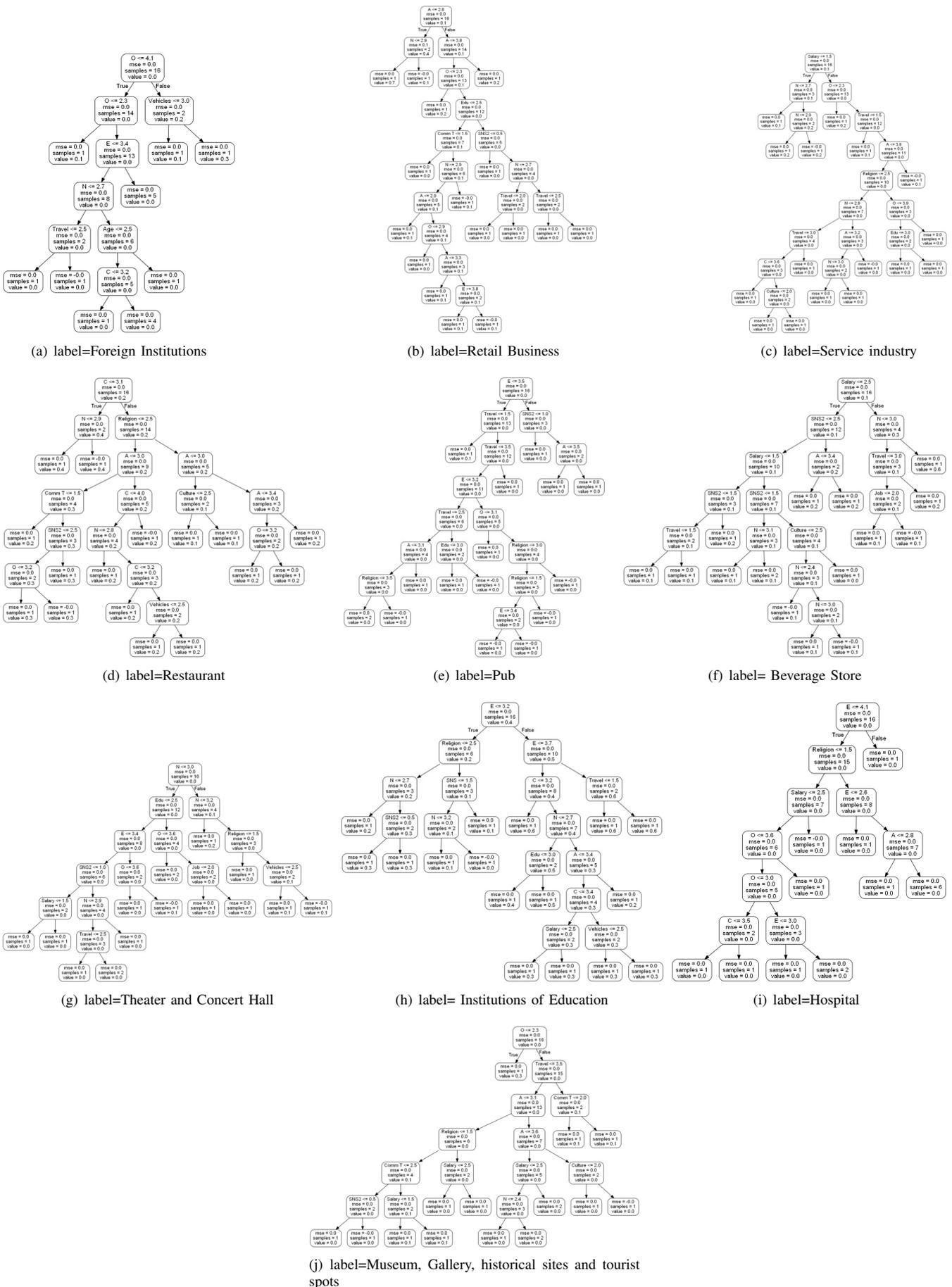


Fig. 1: Decision Tree

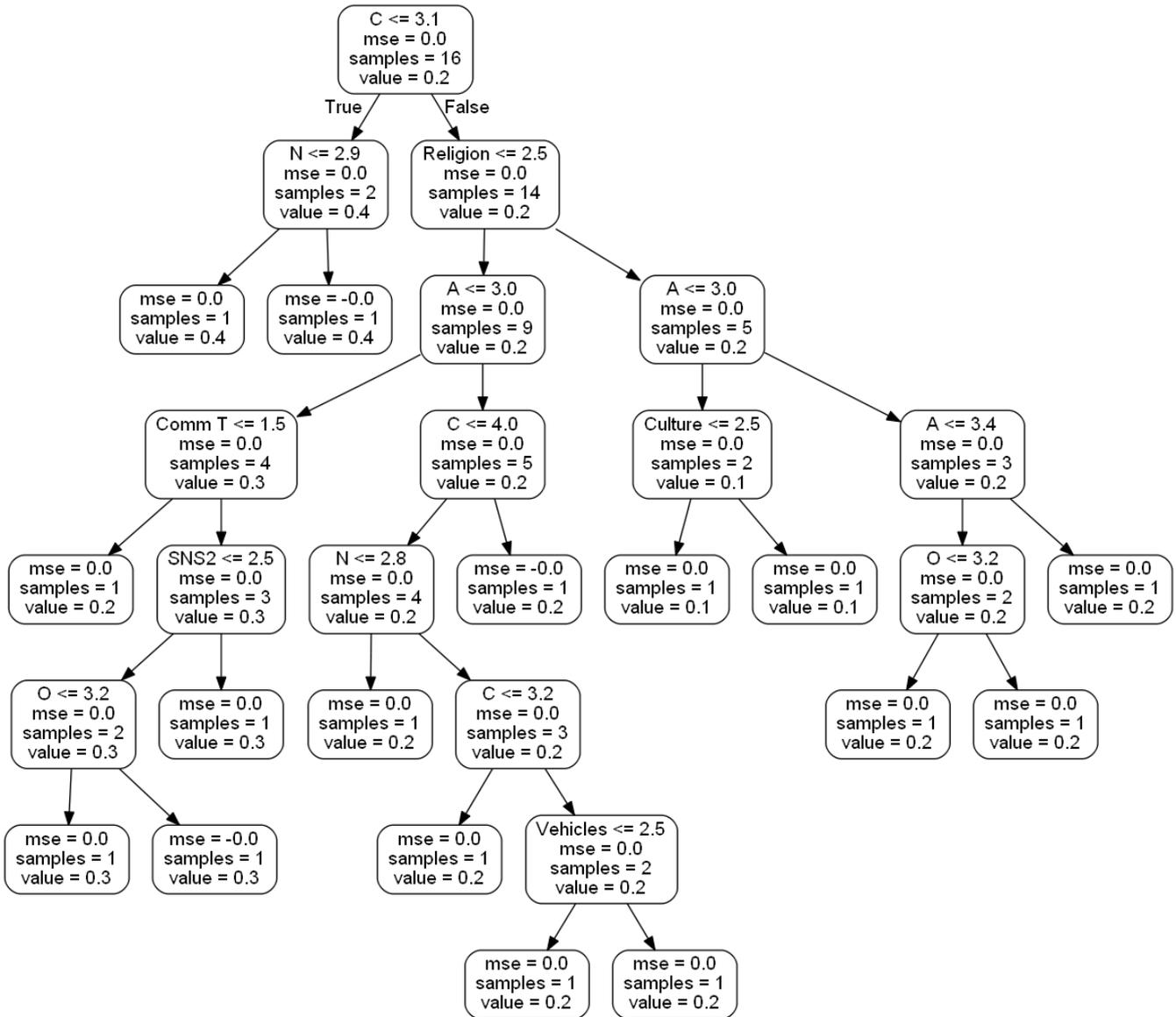


Fig. 2: Decision Tree for Restaurant

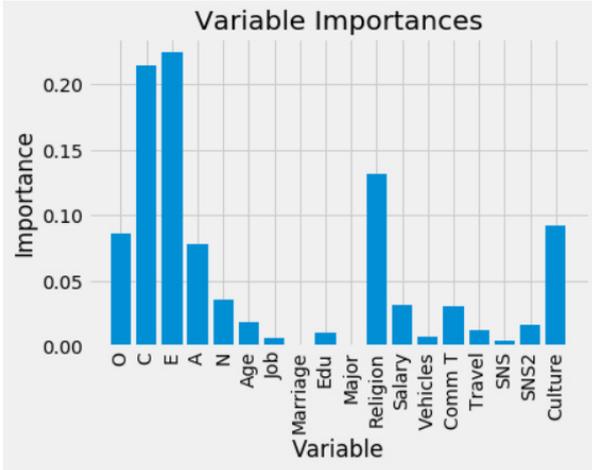
personality or preference. In general, once someone had disease or accident, visit to the hospital will be taken.

For these reasons, most participants rarely went to hospital. One participant frequently visited the hospital during the data collection period because of the need for continuous processing, and this was caused by accident but not by personality or other factors. For foreign institutions, we think significant results could be obtained if the number of participants increased and the age group varied.

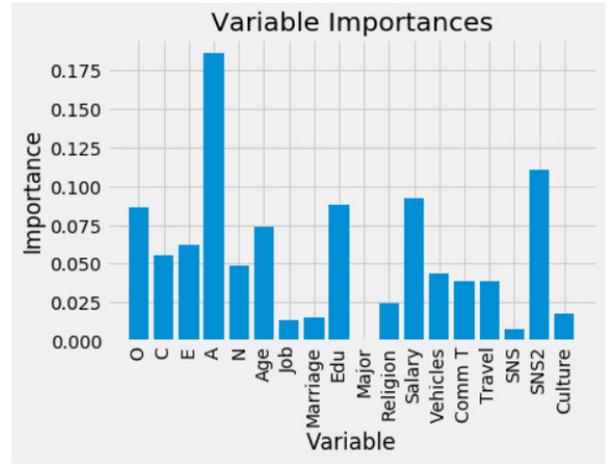
However, in the case of hospital, we decided that the visit frequency was not affected by personality or the personal factors we collected. In the case of service industry, it is difficult to describe this category as specific place because it contains too many locations as previously discussed. For

example, banks, beauty salons, massage parlors, bus terminals, hotels, guest houses and photo studios are included in service industry. These diversities of location category maybe attenuate the accuracy. Prediction accuracy is 59.79%, not very low, but it is also not that high. This would identify better predict accuracy and the affecting factors if the categories were more granular and grouped into units with one characteristic. For the category of museum, gallery, historical sites and tourist spots, the value importance is considered to have a significant result, although the predict accuracy 44.44% which was not high enough.

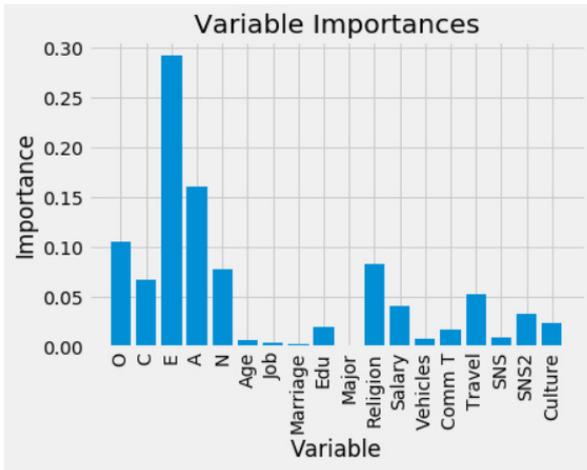
The results of experiments showed that openness and travel frequency affected the visit of museum, gallery, historical sites and tourist spots. Intuitively, open people like to travel because



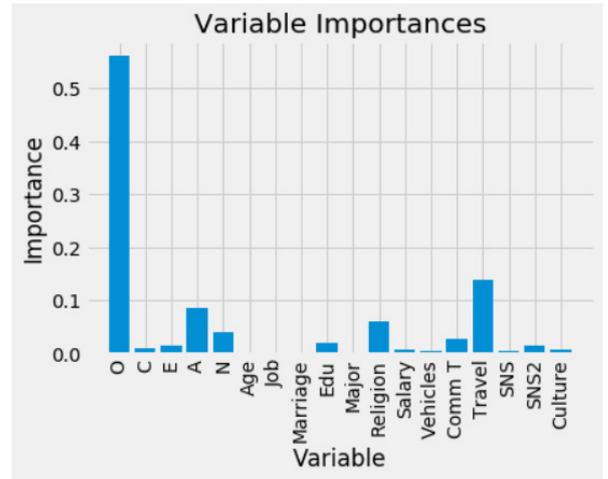
(a) label=Restaurant



(b) label=Pub



(c) label= Institutions of Education



(d) label=Museum, Gallery, historical sites and tourist spots

Fig. 3: Value Importance Graph

TABLE V: Summary of Results

	FI	RB	SI	Restaurant	Pub	BS	TC	IE	Hospital	MG
SMAPE(%)	97.43	50.92	40.21	21.74	34.41	29.97	35.54	23.02	87.4	55.56
Accuracy(%)	2.57	49.08	59.79	78.26	65.59	70.03	64.46	76.98	12.6	44.44
Feature 1	O	A	A	E	A	Religion	N	E	Age	O
	0.55	0.44	0.16	0.22	0.19	0.21	0.21	0.29	0.17	0.56
Feature 2	C	N	Age	C	SNS2	E	Religion	A	Edu	Travel
	0.11	0.12	0.15	0.21	0.11	0.11	0.21	0.16	0.17	0.14
Feature 3	A	O	O	Religion	O	SNS2	Salary	O	A	A
	0.1	0.09	0.12	0.13	0.09	0.09	0.13	0.11	0.13	0.09
Feature 4	E	E	C	O	Edu	N	E	N	E	Religion
	0.06	0.06	0.1	0.09	0.09	0.08	0.08	0.08	0.09	0.06
Feature 5	Job	C	Marriage	Culture	Salary	Salary	C	Religion	C	N
	0.03	0.05	0.09	0.09	0.09	0.08	0.07	0.08	0.08	0.04

TABLE VI: Statistics on Survey

Answers	Age	Job	Marriage	Edu	Major	Religion	Salary	Vehicles	Comm T	Travel	SNS1	SNS2	Culture
1	0	32	1	0	0	23	11	9	13	9	25	4	10
2	30	0	33	25	0	5	17	1	8	16	9	14	8
3	3	2		5	0	3	3	0	13	6		7	16
4	1			3	34	3	1	24	0	3			
5							2						

they love adventure. Frequent travel increases the chances of visiting museum, gallery, historical sites and tourist spots. We also expect to have a high degree of predict accuracy if it gets data from a wider range of ages and occupational groups.

### B. Interpretation of Results

The experimental results show that the predict accuracy is usually high when the characteristics of the category are clear. For example, Restaurant, Pub, Beverage Store, Theater and Concert Hall, Institutions of Education are clear categories. While, Retail Business, Service industry, Museum, Gallery historical sites and tourist spots are not easy to clarify.

Therefore, it is hard to say that the category has one characteristic. For these reasons, it would have been difficult to predict by personality or personal factors. The highest predict accuracy was restaurant category, which was 78.26%. The most affected features are E (Extraversion) and C (Conscientiousness) among personality factors, and followed by Religion, O (Openness) and Cultural Life. For the institutions of education, the predict accuracy is 76.98%, followed by restaurant with a higher predict accuracy. Effective features include E(Extraversion), A(Agreeableness), O(Openness), N(Neuroticism), and Religion.

For the two categories of restaurant and school, we found distinguished results. At this time, it was determined that effective value importance value is greater than 0.1. Considering that most experimental participants of the study are students in their twenties, Extraversion, Agreeableness, and Openness leads to frequent visit to schools. Extraversion, Conscientiousness, and religion also affect the frequent visit to restaurant. To infer why these results came out, we expect that extroverted, enthusiastic and sincere students would have often eaten outside because they would often come to school and stay for a long. Otherwise extroverts are expected to engage in various activities. There would have been many visits to Restaurant in the process. In this context, visits to the beverage store will also have an impact on Extraversion.

Some of the questions in the experiment are that religion has a lot of impact on visiting the beverage store, and the theater and concert hall. In addition to religion, Neuroticism and salary affect theater and concert hall visits.

As mentioned earlier, people with high Neuroticism are stress sensitive and impulsive. Therefore, it is expected that stress will be solved through cultural life such as movie. Also, because cultural life costs, salary will also be effective. For the category pub, Agreeableness, SNS usage frequency, and

Openness are effective. It can be inferred that people who get along well with many people are cooperative, have a communal personality, and often have drinking parties.

### V. CONCLUSION AND FUTURE WORK

In this research, we found that various factors including personality factors effects the selection of visiting place. Especially, factors such as salary, religion, SNS usage were newly distinguished as effective factors for favorite location selection. Several matters must be considered for more precise evaluation. First, most of participants were in their twenties. Table VI shows that several values are skewed. Therefore, these skewed values attenuate the relationship toward visiting places. Once we can get more personal factors including more various age, we guess that more general results with more credible results can be analyzed. Second, we need to adjust location category. For the current categories of location service, two categories contain too many location subcategories. For example, large general retailing and service business contain restaurant and bar but such categorization cannot characterize the locations. This phenomenon leads to inaccurate prediction result. Therefore, ramified categories must be applied in such case so to improve accuracy of analysis. Third, the more data must be collected, especially the location data. Most of participants are not eager to collect their visiting location using SWARM app or does not know the usage of SWARM app. This sort of collection is called as 'check-in'. Collecting continuous geo-positioning data is passive, meaning that the geo-positioning data is automatically collected, while active check-in is required to use SWARM app. For the next research, we need to give more guidance of SWARM to participants. As well, some participants are too eager to collect check-in data so that even bus stops were checked in. This phenomenon may affect the analysis results. Several location categories are regarded as non-associated with personal factors we designed. For example, in case of hospital, accidents or disease may leads to the visit to hospital rather than personal factors. Therefore, personality, gender, hobby is regardless of such locations. Since most of the participants are students, educational locations are frequently visited. Maybe the job of students will affect the visit to educational locations. Therefore, we need to collect more various data to deduce meaningful result. Our analysis result could be applied to various area requiring visiting places prediction. For example, Location Based Service (LBS) and recommendation system

maybe best application area of our research. With the combination of personal factors and favorite visiting places, the usefulness of LBS and recommendation system can have more value added results and high quality of service.

#### REFERENCES

- [1] H. Y. Song and E. B. Lee, "An analysis of the relationship between human personality and favored location," *AFIN* 2015, p. 12, 2015.
- [2] H. Y. Song and H. B. Kang, "Analysis of relationship between personality and favorite places with poisson regression analysis," *ITM Web of Conferences*, vol. 16, p. 02001, 2018. doi: 10.1051/itmconf/20181602001. [Online]. Available: <https://doi.org/10.1051/itmconf/20181602001>
- [3] S. Y. Kim and H. Y. Song, "Predicting human location based on human personality," in *Lecture Notes in Computer Science*. Springer International Publishing, 2014, pp. 70–81. [Online]. Available: [https://doi.org/10.1007/978-3-319-10353-2\\_7](https://doi.org/10.1007/978-3-319-10353-2_7)
- [4] S. Kim and H. Song, "Determination coefficient analysis between personality and location using regression," in *International conference on sciences, engineering and technology innovations*. Bali, ICSETI, 2015, pp. 265–274.
- [5] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001. doi: 10.1023/A:1010933404324
- [6] P. T. Costa and R. R. McCrae, "Four ways five factors are basic," *Personality and Individual Differences*, vol. 13, no. 6, pp. 653–665, jun 1992. doi: 10.1016/0191-8869(92)90236-i. [Online]. Available: [https://doi.org/10.1016/0191-8869\(92\)90236-i](https://doi.org/10.1016/0191-8869(92)90236-i)
- [7] J. Hoseinifar, M. M. Siedkalan, S. R. Zirak, M. Nowrozi, A. Shaker, E. Meamar, and E. Ghaderi, "An investigation of the relation between creativity and five factors of personality in students," *Procedia - Social and Behavioral Sciences*, vol. 30, pp. 2037–2041, 2011. doi: 10.1016/j.sbspro.2011.10.394. [Online]. Available: <https://doi.org/10.1016/j.sbspro.2011.10.394>
- [8] D. Jani, J.-H. Jang, and Y.-H. Hwang, "Big five factors of personality and tourists' internet search behavior," *Asia Pacific Journal of Tourism Research*, vol. 19, no. 5, pp. 600–615, 2014. doi: 10.1080/10941665.2013.773922
- [9] D. Jani and H. Han, "Personality, social comparison, consumption emotions, satisfaction, and behavioral intentions," *International Journal of Contemporary Hospitality Management*, vol. 25, no. 7, pp. 970–993, sep 2013. doi: 10.1108/ijchm-10-2012-0183. [Online]. Available: <https://doi.org/10.1108/ijchm-10-2012-0183>
- [10] O. P. John, S. Srivastava et al., "The big five trait taxonomy: History, measurement, and theoretical perspectives," *Handbook of personality: Theory and research*, vol. 2, no. 1999, pp. 102–138, 1999.
- [11] Y. Amichai-Hamburger and G. Vinitzky, "Social network use and personality," *Computers in Human Behavior*, vol. 26, no. 6, pp. 1289–1295, nov 2010. doi: 10.1016/j.chb.2010.03.018. [Online]. Available: <https://doi.org/10.1016/j.chb.2010.03.018>
- [12] M. J. Chorley, R. M. Whitaker, and S. M. Allen, "Personality and location-based social networks," *Computers in Human Behavior*, vol. 46, pp. 45–56, 2015. doi: 10.1016/j.chb.2014.12.038
- [13] Foursquare Labs, Inc., "Swarm app," <https://www.swarmapp.com/>, 2019.
- [14] G. Biau and E. Scornet, "A random forest guided tour," *TEST*, vol. 25, no. 2, pp. 197–227, apr 2016. doi: 10.1007/s11749-016-0481-7. [Online]. Available: <https://doi.org/10.1007/s11749-016-0481-7>
- [15] M. R. Segal, "Machine learning benchmarks and random forest regression," 2004. [Online]. Available: <https://escholarship.org/uc/item/35x3v9t4>
- [16] International Standard Classification of Occupation, "ISCO," <https://www.ilo.org/>.

# Automated Generation of Business Process Models using Constraint Logic Programming in Python

Tymoteusz Paszun, Piotr Wiśniewski\*, Krzysztof Kluza\* Antoni Ligęza

AGH University of Science and Technology  
al. A. Mickiewicza 30, 30-059 Krakow, Poland

\*E-mail: {wpiotr,kluza}@agh.edu.pl

**Abstract**—High complexity of business processes in real-life organizations is a constantly rising issue. In consequence, modeling a workflow is a challenge for process stakeholders. Yet, to facilitate this task, new methods can be implemented to automate the phase of process design. As a main contribution of this paper, we propose an approach to generate process models based on activities performed by the participants, where the exact order of execution does not need to be specified. Nevertheless, the goal of our method is to generate artificial workflow traces of a process using Constraint Programming and a set of predefined rules. As a final step, the approach was implemented as a dedicated tool and evaluated on a set of test examples that prove that our method is capable of creating correct process models.

**Index Terms**—business process management, process composition, workflow logs, constraint programming, BPMN

## I. INTRODUCTION

THE purpose of the existence of any organization or company is to carry out its mission effectively and efficiently. Lack of coordination of operational activities may result in the ineffective achievement of goals, and in extreme cases may lead to failure of the entire undertaking. This is a particular threat to enterprises, understood here as organizations whose mission is to provide specific products in the form of goods or services. In their case, permanent failure to meet the clients' needs usually results in bankruptcy or severe difficulties in operating in a competitive market. In order to minimize the risk of such turnover, the activities carried out within the organization in the form of processes are often created. As the processes are often complex, their modeling is a challenge for business analysts. To facilitate this task, it is possible to use some tools to assist analysts in their daily work. This paper combines issues in the areas of management (process approach in organizations) and Information Technology (Constraint Programming, Process Mining).

The main goal of the approach is to provide a method to generate complex process models starting from tasks and constraints obtained from the organization. The construction of the solution was preceded by the analysis of this topic.

This paper is organized as follows: Section II presents an analysis of the Business Process Management approach, including its origins, development, and current trends. The section also includes necessary information related to BPMN and process discovery. In Section III, as the next step to achieving the goal, the analysis and description of the proposed method are included. Next, a project of an IT tool, its assumptions,

requirements, and architecture (Section IV) was presented, as well as the technical description of its implementation (Section V). Section VI includes the evaluation of the proposed approach, the description of the developed tool, as well as the results of its application on a set of test data. The work is finished with conclusions and a description of the possible extension of the approach (Section VII).

## II. BUSINESS PROCESS MANAGEMENT

This Section discusses the issues related to business processes – from their role in management, through the applied notation of their recording, to the description of their research techniques.

### A. Overview

Business Process Management (BPM) [1] is one of the most common methods for improving the organization and implementation of the quality system. The ISO 9001 standard [2] introduced the obligation to apply the process approach as one of the key elements of a well-implemented, maintained, and functioning management system.

However, to talk about the process approach, let us look at the concept of a business process per se. There are many definitions of the business process. However, for this study, the definition presented in "Essential Business Process Modeling" will be adopted, where the process is described as step-by-step activities specific to the solution of various problems or business issues [3].

To answer the key question about the purpose of the process approach in enterprises, it is worth taking a closer look at the research carried out in 2009 [4], which shows the organization's goals at various stages of process development. Enterprises in the phase of introducing process-oriented approach set the implementation of the quality system and the creation of a process approach in their structures as the primary goal. In the case of enterprises from this group, the aim is usually to map processes existing in the organization, and less frequently to improve their effectiveness or implement IT tools supporting operational activities. For 38.5% of organizations in the growth phase, the key goal is to improve efficiency, and for 31% of organizations in this group, the most important goal is the development of applied IT systems, which will also improve the organization's efficiency [4]. Enterprises being in the improvement phase during the research indicated three

most important areas of application of the process approach, which are: improvement of quality, improvement of efficiency, and implementation of IT systems [4].

In the mid-nineties, Business Process Management was introduced as the next wave of approach to managing the processes in the organization. BPM postulates, inter alia, mapping, visualization, and analysis of processes in the organization. Thanks to these activities, it is possible to standardize the implemented activities, control their course, and perform much easier analysis of decision situations [5], [6].

Modeling processes in firms can be implemented using a variety of notations. Initially, many organizations used their own methods of describing and modeling processes, which, however, hindered readability and negatively affected cooperation between organizations. In response to this problem, many formal notations and languages of business process modeling were created. The most popular standards used to model business processes are the Unified Modeling Language (UML) [7] and Business Process Model and Notation (BPMN) [8].

In the last three decades, a change in the approach to information systems can be observed. Process-aware systems increasingly replace formerly used data-aware systems. To support business processes implemented within the framework of an organization, enterprise information systems must be somewhat aware of the existence of these processes and the organizational context within which they are implemented. Early examples of process-aware systems were called Workflow Management systems (WFM). In recent years, IT solutions companies have preferred a more precise term of BPM. Business Process Management systems cover a wider range than classical Workflow Management systems and do not focus only on process automation. Business Process Management systems attach more importance to supporting various forms of analysis (e.g. process simulation) and management (e.g. monitoring key performance indicators). Both Workflow Management systems and Business Process Management systems seek to support operational processes, which we refer to as workflow processes or simply workflows [9].

### B. Business Process Model and Notation

The BPMN standard consists of a set of graphic elements for constructing diagrams showing components of the process and the way in which it should be executed. Graphic symbols and the way of combining them constitute the syntax of the notation and have defined semantics. In addition to the graphical representation of the model, the BPMN specification [10] also describes the format of the record in the form of XML files.

The basic subset of BPMN symbols used in the developed model generator is presented in Figure 1, its elements are described below.

- Task – atomic activity performed as part of the process which is not subject to further decomposition. It presents actions taken by the end user or software.
- Gateways – elements used to control how sequential flows separate and merge as part of the process. They can

support many input and many output flows, although best practices suggest that the gateway should only perform one of these functions. Therefore, in the diagrams, the pair of gates usually serves first to separate and then connect the process flows.

- OR-gateway – flow object used to create alternative paths within the process flow. A single process instance contains only one of the possible paths selected. This gateway is interpreted as a decision at a given point in the process. It can be understood as a question, and sequence flows from it will be associated with responses.
- AND-gateway – flow object used to create parallel flows or synchronize them (join). Separation of the flow is not subject to any conditions, the connection, in turn, requires the completion of all input flows of the gate. A single process instance contains all possible paths associated with a given pair of parallel gates.
- Process start and end event – process start event indicates where flows begin within a given process. The process end event symbolizes the end of all flows.
- Sequence flow – indicates the order of flow objects in the process. It always has one source and one target element.

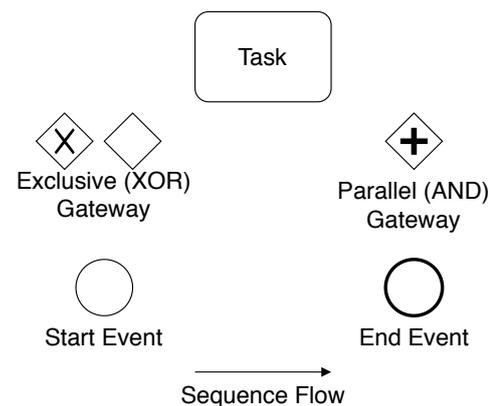


Figure 1. BPMN elements.

### C. Process Discovery

Process mining is a relatively new field of research, between machine learning and data exploration on the one hand, and modeling and analysis of processes on the other. Today's information systems store huge amounts of data about activities performed in the form of event logs. The assumptions for the exploration of processes are to discover, control, and streamline real processes by extracting knowledge from read-only event logs in these systems.

In general, process mining methods are often classified into one of the following classes [11]:

- process discovery,
- conformance checking,
- process enhancement.

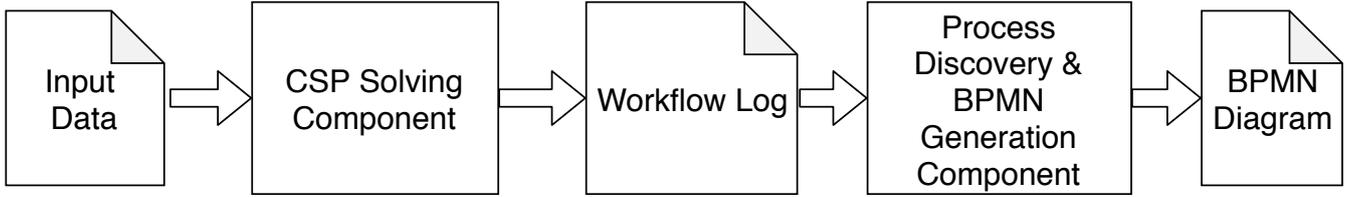


Figure 2. Overview of the process generation method.

For the purpose of this paper, our attention will be focused only on the first group – methods of discovering the process model. Discovery techniques rely on event logs and generate a process model without any known process information. An example is the  $\alpha$  algorithm, which results in a Petri net [12] that reflects the behavior recorded in the event log. Using dedicated algorithms, Petri nets can be automatically converted to BPMN [13], [14], [15].

Example algorithms of process discovery include:

- $\alpha$  algorithm [16], used in the developed tool,
- Inductive Miner [17], used in the developed tool,
- Heuristics Miner [18] and its Fodina variant [19],
- Evolutionary Tree Miner [20],
- Structured Miner [21],
- Region-Based Mining [22],
- Split Miner [23].

### III. METHOD

The process model generator presented in this paper uses the adapted method of generating process models based on the one described in the work [24]. It consists of three stages presented in Figure 2. In this Section, the formalization of the method is presented.

The first stage of the method is the preparation and provision of input data in the appropriate format. Such data can be extracted from a system-based source such as data warehouse [25] or acquired directly from business roles [24]. The approach described in this paper requires the provision of input data consisting of:

- matrix for prerequisite tasks,
- matrix of task effects,
- vector of the initial state,
- matrix of acceptable final states of the process,
- the maximum number of executions of each task.

The next stage of the method's operation is the use of a component that uses the Constraint Programming techniques, based on a model built of input data and predefined constraints. The result of its operation is an artificially created log of all possible traces of the process.

The final stage of the method is the part that explores the process (called *process exploration*) directly from the event log and creates a process representation in BPMN.

#### A. Expected Input Data

In our approach,  $\mathbb{T}$  denotes a set of all tasks:

$$\mathbb{T} = \{\tau^{(1)}, \tau^{(2)}, \dots, \tau^{(n)}\}.$$

Table I  
MEANING OF VALUES IN USED STRUCTURES.

Value	$M_{TC}$	$M_{TE}$	$M_{ST}$	$s_0$
-1	not relevant	not changed	not relevant	-
0	forbidden	deleted	forbidden	forbidden
1	required	created	required	required

Tasks, or activities, are performed within the course of the process. For our method, the concept of data entity was introduced. Unlike data objects used in BPMN, data entities do not exist in the generated model, but a part of the process specification that is required in the applied method. In other words, the data entity is a variable about a simple or complex type of data that accompanies the execution of the tasks of the process. The set of all data entities is denoted by  $\Delta$ :

$$\Delta = \{\delta_1, \delta_2, \dots, \delta_m\}.$$

Cardinality of  $\mathbb{T}$  and  $\Delta$  is equal to  $n$  and  $m$ , respectively. For the needs of the constraint model, it is necessary to build two matrices of dimensions  $n \times m$ :

- 1)  $M_{TC}$ : for the prerequisites required for each task,
- 2)  $M_{TE}$ : for the effects caused by each task.

The prerequisites and effects are understood as the occurrence of the data entity before and after the task.

Additionally, assuming  $g$  as the number of allowed final states, it is necessary to define the matrix  $M_{ST}$  of dimensions  $g \times m$ , which describes all acceptable terminal states. The  $m$ -element vector  $s_0$  is defined in order to give information about the presence of the data entity before the process is executed. All structures described can contain integer values from the set  $\{-1, 0, 1\}$ . Table I explains the meaning of values in the context of the data entity in each structure.

The last of the input structures is the  $n$ -element vector  $e_t$  containing the number of maximum executions for each task. By default, its values should be equal to 1 unless the process contains loops or tasks performed iteratively.

#### B. Workflow Log

Workflow log  $W = \{\sigma_1, \sigma_2, \dots, \sigma_L\}$  is a multiset of individual workflow traces  $\sigma$ , which can be defined as ordered sequences of activities in the course of the process:  $\sigma = (\tau_1, \tau_2, \dots, \tau_K), \tau_i \in \mathbb{T}$ . Although the workflow log definition permits the appearance of identical process traces

many times, the purpose of the described method is to generate a complete log artificially. The generated log contains all acceptable process traces. Therefore, in further considerations, the multiset  $W$  will be treated as an ordinary set.

### C. Constraints

For the purpose of finding a set of solutions, the concept of the process state  $S$  is introduced. It is represented by the state vector of the data entity in every step of the process. The state of the data entity  $s_i$  is the vector representing the occurrence of the data entity in  $i$ -th step of the process. The values 0 and 1 mean respectively the absence and occurrence of the data entity.

$$S = [s_0, s_1, \dots, s_K], \text{ where } K > 0, K \in \mathbb{N},$$

$$s_i = [d_{(i,1)}, d_{(i,2)}, \dots, d_{(i,m)}],$$

$$d_{(i,j)} \in \{0, 1\}, \text{ where } i \in \{1, \dots, K\}, j \in \{1, \dots, m\}.$$

Before specifying the constraints needed to generate the correct process flow log, it is necessary to define a predicate that determines whether the data vector state of the data entity  $s_i$  meets the requirements of the task to be performed:

$$sat(s_i, TC(\tau^{(i)})) \iff$$

$$\forall j = 1 \dots m : d_{(i,j)} = TC(\tau^{(i)})_j \vee TC(\tau^{(i)})_j = -1,$$

where  $TC(\tau^{(i)})$  is the  $i$ -th row of matrix  $M_{TC}$  and  $d_j$  is the  $j$ -th element of state vector  $s_i$ .

In addition, a predicate is defined, meaning that the state meets one of the allowed end states:

$$satSet(s_i, M_{ST}) \iff \exists j = 1 \dots g : sat(s_i, M_{ST_j}),$$

where  $M_{ST_j}$  means the  $j$ -th row of admissible solution matrix.

To generate a complete workflow log  $W$ , the problem being analyzed must be modeled using constraints over variables. This concept is based on three principles:

- 1) Search space: all completed task sequences.
- 2) Decision variables: single process flow, process state matrix.
- 3) Variable constraints: defined by the input data as well as by the set of predefined rules.

Predefined constraints that ensure the correctness of the generated process runs are:

- 1) The overall limit of task executions  $MAX_{EX}$ .
- 2) The number of executions of each  $\tau^{(i)}$  task must be less than or equal to the corresponding value in the vector of the maximum number of executions of the  $e_t$  task or the general  $MAX_{EX}$  limit.
- 3) Maximal length of a single workflow trace  $\sigma$  to  $K = n \times MAX_{EX}$ .
- 4) The input state of the first completed task is equal to  $s_0$ .
- 5) Every non-idle task  $\tau^{(k)}$  in the  $i$ -th step changes the elements of its successor state  $s_{i+1}$ :

$$s_{i+1} = [d_{(i+1,1)}, d_{(i+1,2)}, \dots, d_{(i+1,n)}],$$

$$d_{(i+1,j)} = \begin{cases} d_{(i,j)}, & \text{dla } TE(\tau^{(k)})_j = -1, \\ 1, & \text{dla } TE(\tau^{(k)})_j = 1, \\ 0, & \text{dla } TE(\tau^{(k)})_j = 0, \end{cases}$$

where  $TE(\tau^{(k)})_j$  means the  $j$ -th element of the  $k$ -th row of matrix  $M_{TE}$ .

- 6) The process ends when one of the specified final states is reached.

$$satSet(s_i, M_{ST}) \iff \tau_i = \tau^{(0)},$$

where  $\tau_i$  means the  $i$ -th task of a single workflow trace  $\sigma$  and  $\tau^{(0)}$  is an idle task.

- 7) The last process state  $s_K$  satisfies one of the admissible goal states  $M_{ST}$ :

$$satSet(s_k, M_{ST}).$$

- 8) The task can be performed only if the current state meets its initial conditions:

$$\tau_i = \tau^{(k)} \iff sat(s_i, TC(\tau^{(k)})).$$

The program built on the basis of the above-mentioned constraints and input data is performed by the system solving the problems of meeting restrictions (called *solver*). Executing a program by a solver to find all solutions results in an artificially generated process log  $W$ , needed in the next stage of generating the process model.

### D. Generating a BPMN model from a workflow log

In the work [24], two approaches to building a process model based on the delivered process log are listed. The first approach involves the use of process discovery algorithms from the delivered process log. These algorithms were described in more detail in Section II-C. The result of their operation is the process model in the form of a Petri net. One of the methods of converting Petri nets to BPMN was presented in [14]. The second approach is a process composition method based on activity graphs that does not require conversion of the Petri net to the BPMN form because the BPMN composition is directly from the artificially generated workflow log.

In the process model generator described in this paper, the implemented implementations of process discovery and conversion algorithms for BPMN contained in library *Numberjack* were used. The tool uses the  $\alpha$  and Inductive-Miner algorithms described in [26] and [17] respectively.

## IV. TOOL

This section presents the tool design – from specifying functional and non-functional requirements, by specifying the input data format specification, to the architecture description of the developed process model generator.

At the initial stages of work, functional (Section IV-A) and non-functional requirements (Section IV-B) were defined.

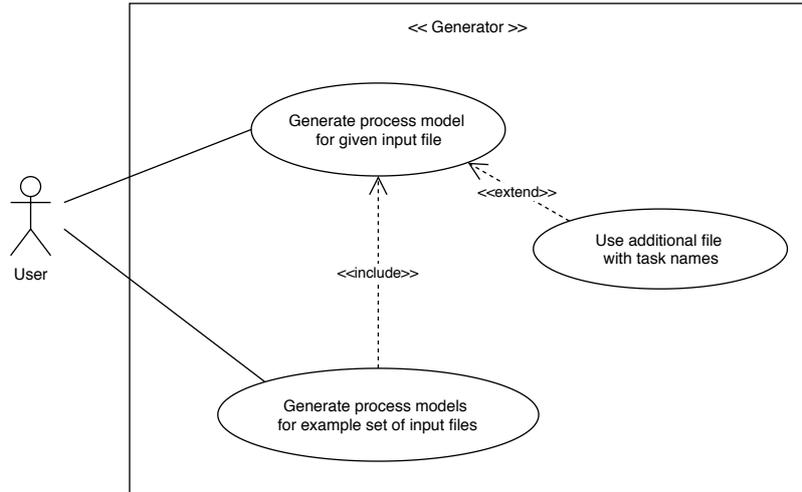


Figure 3. Use case diagram.

A. Functional Requirements

- 1) The tool should accept a set of input data in a specified format at the input.
- 2) The tool should optionally accept a file with the specification of task names at the input.
- 3) The tool as a result of the action should generate:
  - a) BPMN diagrams in graphic form,
  - b) BPMN diagrams in XML format according to the standard,
  - c) artificially generated process log,
  - d) Petri net diagrams resulting from the operation of process discovery algorithms.

Functional requirements are presented in the use case diagram (Figure 3).

B. Non-functional Requirements

- 1) The tool is distributed in an easy-to-use form, in particular without having to manually install all dependent libraries.
- 2) The tool is implemented with division into independent modules so that it can be further expanded.

C. Input File Format

The input tool accepts text files with the input data value definition described in Section III-A. The file should contain in sequence:

- initial state vector,
- matrix of task prerequisites,
- task results matrix,
- matrix of allowed final states,
- vector of the possible number of task executions.

The vectors and matrices should be separated from each other by at least one empty line. The line in the file may contain a comment beginning with the # character. Comments are ignored by the input data parser. The values of subsequent elements of the introduced vectors and matrices are separated by commas. An example input file is shown in Listing 1.

Listing 1. Example input file of the tool.

```
# s_0
0, 1, 0, 0

# m_tc
0, 1, 0, 0
0, 1, 1, -1
0, 1, 1, 0

# m_te
-1, -1, 1, -1
1, -1, -1, -1
-1, -1, -1, 1

# m_st
1, -1, -1, -1

# e_t
2, 2, 2
```

D. Architecture

While working on the tool, three functional areas were identified as separate modules. The modular division of the tool is aimed at introducing a clear structure of responsibility for individual parts, as well as facilitating further development by providing other implementations. The identified software modules are:

- 1) Parser module for input files.
- 2) Log module of the process log generator (based on programming techniques with limitations).
- 3) Module for process discovery and generation of process model diagrams.

The modules are presented in the component diagram (Figure 4). Coordination of the use of these modules by the tool is presented in the sequence diagram (Figure 5).

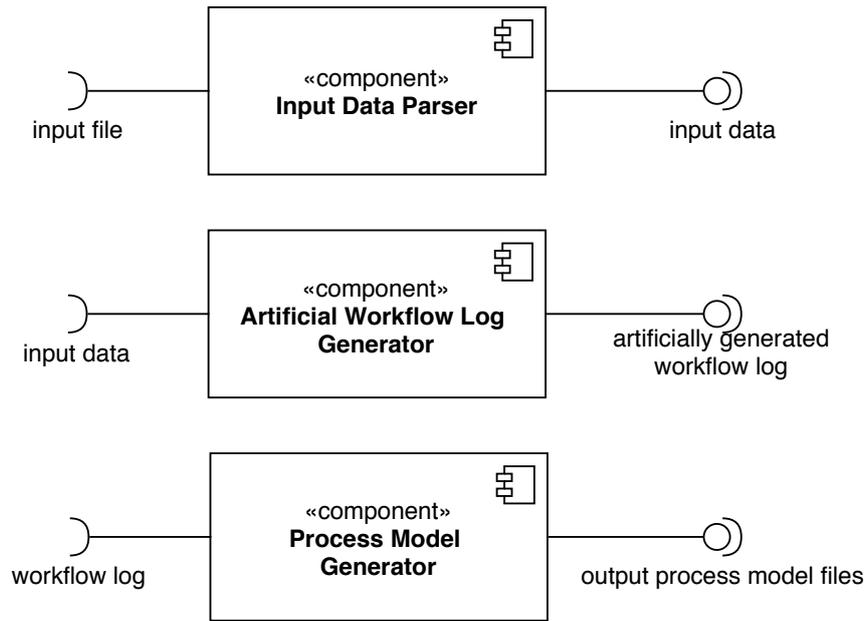


Figure 4. Components diagram.

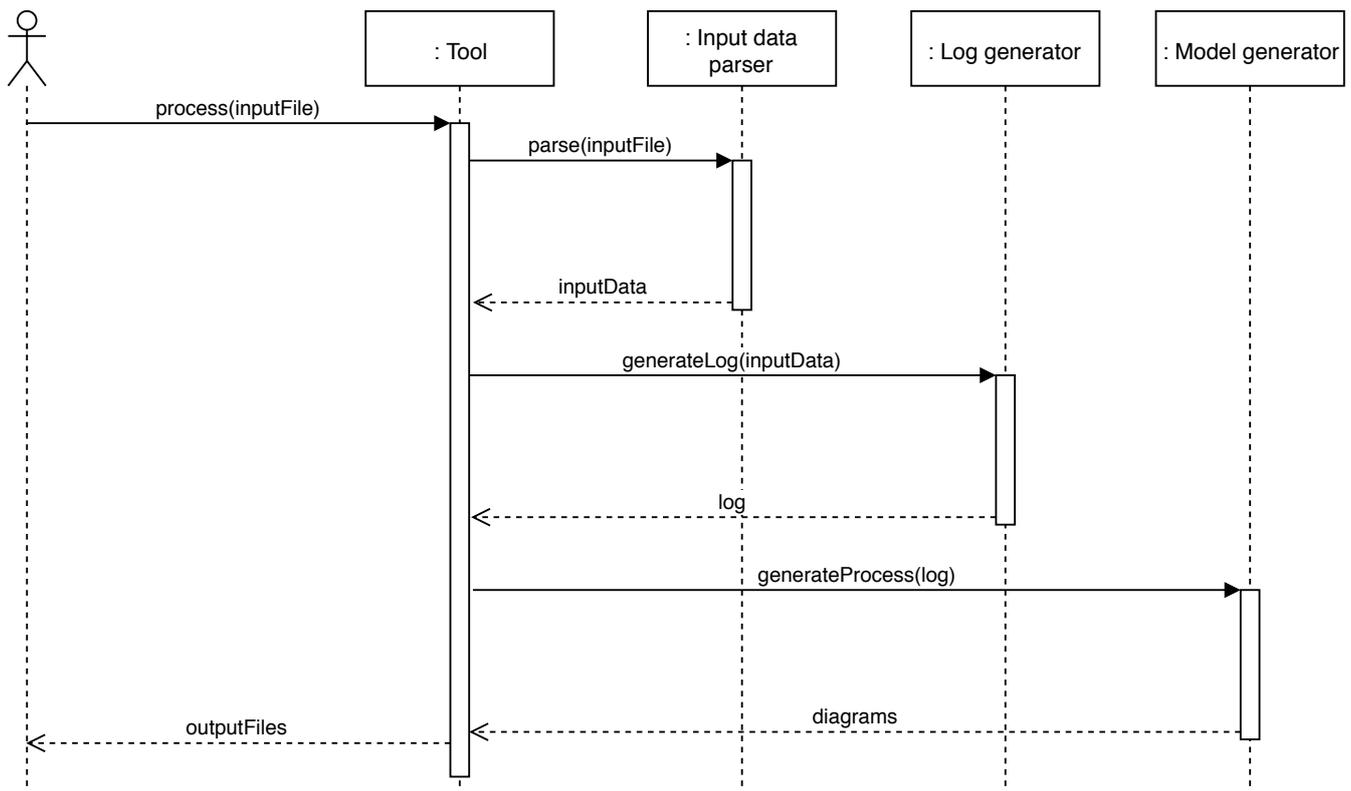


Figure 5. Sequence diagram of model generation.

V. IMPLEMENTATION

In this section, we present the details of the implementation of the process model generator.

A. Constraint Programming

Constraint Programming is a technique for solving the problems of satisfying constraints. These problems can be defined using variables that take values from their domains and the constraints over variables. The solution to the problem is a set of variable value assignments that meet the given constraints [27].

In addition, the problems expressed in Constraint Programming languages are characterized by declarativeness, i.e. a description of the problem becomes a program solving this problem [28].

The Constraint Programming technique can be used in imperative languages by means of libraries that allow building a problem model (variables and constraints) with the help of structures appropriate to a given language. These libraries can themselves implement solvers or provide an interface to solvers implemented in other languages. Examples of widely used solvers are: CP-SAT Solver from the Google OR-Tools package, Gecode, Mistral, Mistral2, ILOG Solver.

B. Used Techniques

Python was chosen as the implementation language of the developed tool. It is a popular language in the academic environment, as well as widely used in the IT industry. Support for programming with restrictions is ensured by the library *Numberjack* (<https://github.com/eomahony/Numberjack>), which allows for high-level modeling of problems and the use of several solvers. Discovering the process is carried out using the library *pm4py* (<http://pm4py.pads.rwth-aachen.de/>), which provides the implementation of the algorithms: *alpha* and *Inductive Miner*. It enables the presentation of discovered processes in the form of Petri nets, as well as their conversion to the BPMN diagram in XML and graphical format. Functions related to BPMN diagram support are currently in the implementation phase (they are not shared with a stable version of the library) and have been taken from the appropriate branch of the library version control repository. The result of their use is described in more detail in the Section VII.

For the purpose of easily recreating the entire working environment of the tool and simplifying its use on other computers, the *Docker* software (<https://www.docker.com/>) used for virtualization at the operating system level was used. The application image, containing the generator of process models and the configured environment for its launch and proper operation, was prepared.

VI. EVALUATION

As part of the research, tests were carried out on synthetic examples. Simple processes have been selected that contain:

- a single task,
- sequential submission of two tasks,
- two tasks covered by the XOR-gateway,

- d two tasks performed concurrently,
- e two tasks covered by XOR-gateway preceded and completed by one task,
- f two tasks performed concurrently preceded and ended with one task,
- g three tasks performed sequentially, in which the middle one is optional.

They are presented in Figure 6. For the needs of the tests, input data corresponding to these processes was prepared.

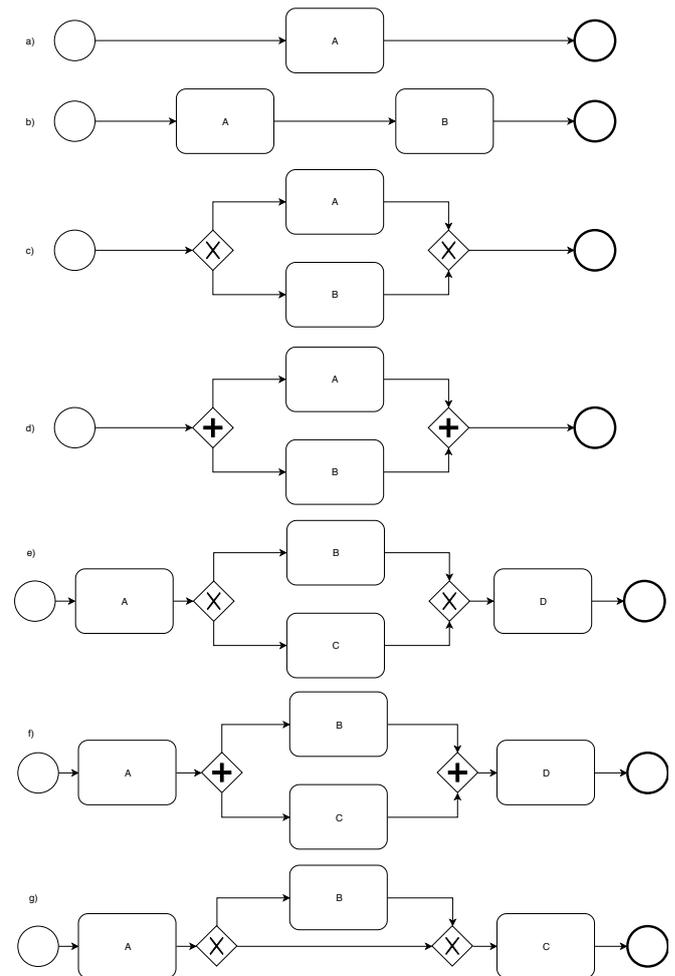


Figure 6. Example test cases in BPMN.

Analysis of test results was divided into two parts. The first one included verification of the correctness of the generated process logs. The second one focused on checking the generated model diagrams.

A. Generated Workflow Logs

The first of the test models – a process containing a single task – has only one possible trace consisting of this task. Also, the second test model, which is the composition of two tasks, should generate one pass. The generated process logs (respectively "A" and "A" "B" confirm the correctness of this stage of the tool for given models.

In the case of two successive process models (two tasks connected with the OR gate, two tasks performed concurrently), we expect logs consisting of two process runs. The correct behavior of the tool was also found here - logs were created with the forms "A", "B" and "A""B", "B""A".

The next two test cases are the extension of the previous ones by adding tasks at the beginning and end of the process. Here too, the proper generation of logs has been observed ("A""B""D", "A""C""D" and "A""B""C""D", "A""C""B""D").

The last example (optional execution of the task in the middle of the process) gives the tool the expected log in the form "A""B""C", "A""C".

The above analysis confirms the correctness of the method used to generate an artificial process log for each of the test cases.

### B. Generated Process Models

The result of the comparison of generated process model diagrams in the form of BPMN for test cases is shown in Table II.

The following deviations from the expected results were found during the analysis:

- 1) algorithm  $\alpha$  for example c) generated BPMN diagram without XOR-gateways (Fig. 7),
- 2) Inductive Miner for example c) generated a BPMN diagram containing doubled XOR-gateways (Fig 8),
- 3) algorithm  $\alpha$  for example d) generated a BPMN diagram without parallel gateways (Fig. 9),
- 4) algorithm  $\alpha$  for example f) generated a BPMN diagram not containing initial and final tasks (Fig. 10),
- 5) algorithm  $\alpha$  for example g) generated a BPMN diagram without the first task and with the wrong XOR-gateway instead of a parallel one (Fig. 11).

Table II  
GENERATED MODELS AND THE EXPECTED RESULTS.

Test case	$\alpha$ Algorithm	Inductive Miner
single task (a)	correct	correct
two tasks (b)	correct	correct
XOR-gateway (c)	incorrect	incorrect
parallel gateway (d)	incorrect	correct
XOR-gateway in a process (e)	correct	correct
parallel gateway in a process (f)	incorrect	correct
optional task (g)	incorrect	correct

### C. Example

Apart from test cases, the tool was also evaluated on the basis of a more complicated example – the process of opening a bank account. Synthesis of the input file was made on the basis of the process initially presented in [29]. The results of BPMN generation using the Inductive Miner algorithm are shown in Figure 12.

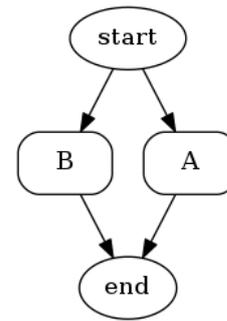


Figure 7. Diagram generated by the  $\alpha$  algorithm for case c)

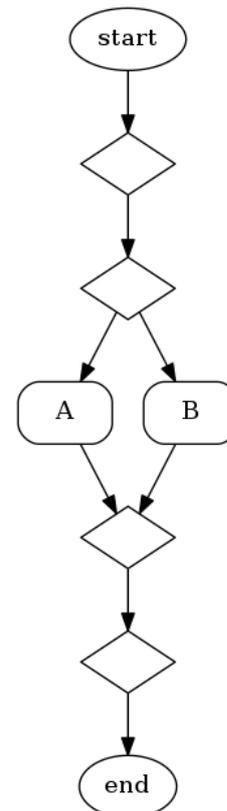


Figure 8. Diagram generated by Inductive Miner for case c)

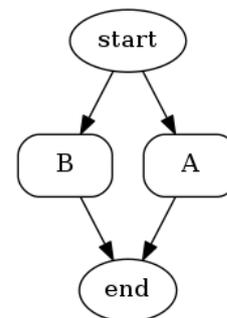


Figure 9. Diagram generated by  $\alpha$  algorithm for case d)

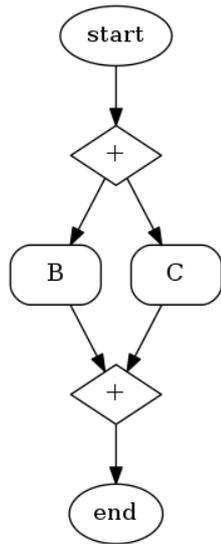


Figure 10. Diagram generated by  $\alpha$  algorithm for case f)

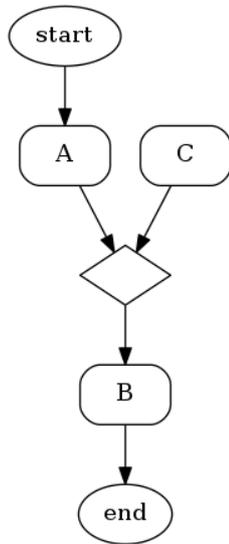


Figure 11. Diagram generated by  $\alpha$  algorithm for case g)

VII. CONCLUSIONS AND FUTURE WORKS

The works on the generator of process models presented in this paper are topped off with half-hearted success. The module for generating an artificial log of the process flow provides the correct results. They can be the basis for discovering the process. However, the process discovery and BPMN diagram generation module generates diagrams far from expectations. Inconsistencies appear at the stage of testing the test cases. If the observed behavior is recorded in an event log, it is possible to repair such a model [30]. However, our goal is to provide a prototype model based on the provided input data. Thus, in order to accurately diagnose the reason for this behavior of the module, a broader review of process discovery algorithms

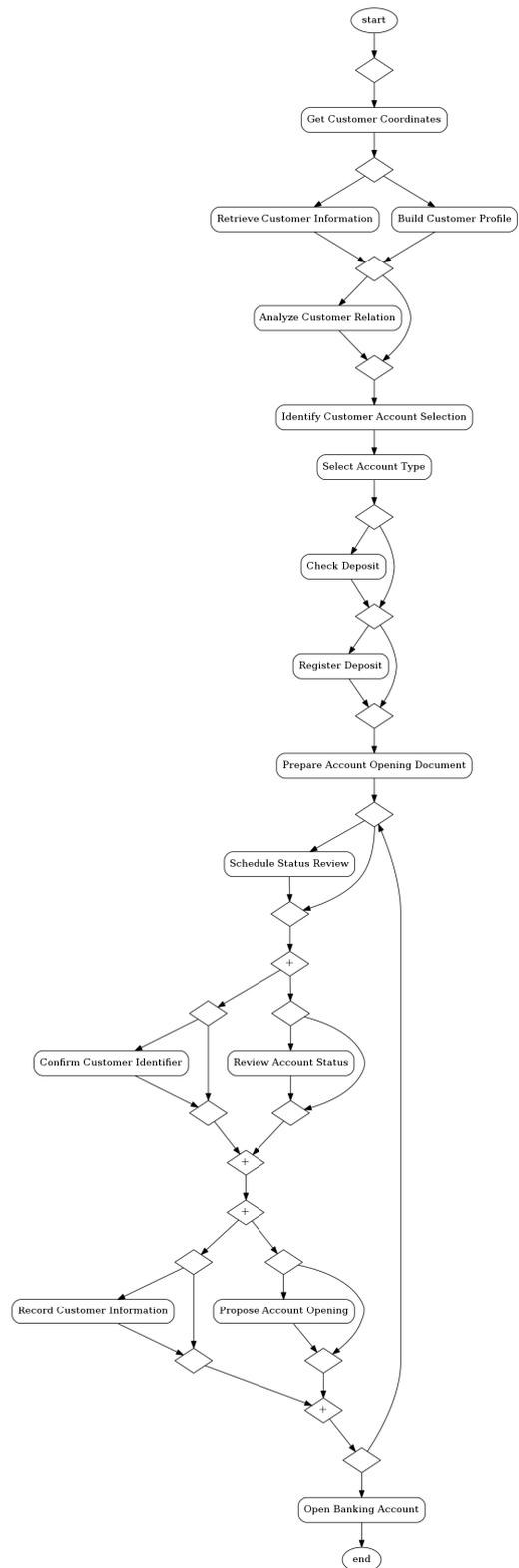


Figure 12. Diagram generated by Inductive Miner for the example process of bank account opening.

and a comparison of their properties should be made. The implementation of the Petri net conversion algorithm to BPMN provided in the experimental branch of *pm3py* can also be one of the causes of the problem.

The possibilities of the process model generator extensions are as follows:

- The use of other process discovery algorithms mentioned in the section II-C, in particular the implementation of the process composition method based on the activity graphs described in work [24]. The use of this approach should give better results during the process discovery phase from the artificially generated log flow of the process than the  $\alpha$  and Inductive-Miner algorithms used.
- Performing a GUI facilitating the input of data, or enabling cooperation with the organization's business roles in order to collect information about tasks, data entities and their relationships.
- Improving the layouting of the generated BPMN diagrams, arranging elements on the diagram in a way similar to how they are visualized in commercial tools.
- Extension of the tool to include information about various departments within the organization, linking them to tasks and extending the generated model with pools and lanes.
- Adding the possibility of exporting the generated process log to the standardized event log format. The XES format is the standard for text-event logs for further analysis using tools that implement the process discovery functions (for example, the ProM framework [31]).

#### REFERENCES

- [1] M. Dumas, M. La Rosa, J. Mendling, H. A. Reijers *et al.*, *Fundamentals of business process management*. Springer, 2013, vol. 1.
- [2] V. Jovanovic and D. Shoemaker, "Iso 9001 standard and software quality improvement," *Benchmarking for Quality Management & Technology*, vol. 4, no. 2, pp. 148–159, 1997.
- [3] M. Havey, *Essential business process modeling*. O'Reilly Media, Inc., 2005.
- [4] W. Ciesliński, "Procesowa orientacja przedsiębiorstw: wyniki badań empirycznych," *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, no. 52 Podejście procesowe w organizacjach, pp. 41–48, 2009.
- [5] S. Lusk, S. Paley, and A. Spanyi, "The evolution of business process management as a professional discipline," *BP Trends*, vol. 20, pp. 1–9, 2005.
- [6] E. Kucharska, "Heuristic method for decision-making in common scheduling problems," *Applied Sciences*, vol. 7, no. 10, p. 1073, 2017.
- [7] K. Kluza, P. Wiśniewski, K. Jobczyk, A. Ligeza, and A. Suchenia (Mroczek), "Comparison of selected modeling notations for process, decision and system modeling," in *Proceedings of the 2016 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems, M. Ganzha, L. Maciaszek, and M. Paprzycki, Eds., vol. 11. IEEE, 2017, pp. 1095–1098.
- [8] P. Pasamonik, "Modelowanie procesów biznesowych zorientowane na czynności," *Zeszyty Naukowe Wyższej Szkoły Informatyki*, vol. 9, no. 2, pp. 102–116, 2010.
- [9] W. M. van der Aalst, "Process-aware information systems: Lessons to be learned from process mining," in *Transactions on petri nets and other models of concurrency II*. Springer, 2009, pp. 1–26.
- [10] OMG. (2011) Business process model and notation. [Online]. Available: <https://www.omg.org/spec/BPMN/2.0>
- [11] W. M. van der Aalst, A. Adriansyah, A. K. A. De Medeiros, F. Arcieri, T. Baier, T. Blickle, J. C. Bose, P. van den Brand, R. Brandtjen, J. Buijs *et al.*, "Process mining manifesto," in *International Conference on Business Process Management*. Springer, 2011, pp. 169–194.
- [12] M. Szpyrka, *Sieci Petriego w modelowaniu i analizie systemów współbieżnych*. Wydawnictwa Naukowo-Techniczne, 2008.
- [13] A. A. Kalenkova, M. De Leoni, and W. M. van der Aalst, "Discovering, analyzing and enhancing BPMN models using ProM," in *BPM (Demos)*, 2014, p. 36.
- [14] A. A. Kalenkova, W. M. van der Aalst, I. A. Lomazova, and V. A. Rubin, "Process mining using BPMN: relating event logs and process models," *Software & Systems Modeling*, vol. 16, no. 4, pp. 1019–1048, 2017.
- [15] A. Kalenkova, A. Burattin, M. de Leoni, W. van der Aalst, and A. Sperduti, "Discovering high-level BPMN process models from event data," *Business Process Management Journal*, 2018.
- [16] W. M. van der Aalst, A. Weijters, and L. Maruster, "Workflow mining: Which processes can be rediscovered," BETA Working Paper Series, WP 74, Eindhoven University of Technology, Eindhoven, Tech. Rep., 2002.
- [17] S. J. Leemans, D. Fahland, and W. M. van der Aalst, "Scalable process discovery with guarantees," in *International Conference on Enterprise, Business-Process and Information Systems Modeling*. Springer, 2015, pp. 85–101.
- [18] A. Weijters and J. Ribeiro, "Flexible heuristics miner (fhm)," in *2011 IEEE symposium on computational intelligence and data mining (CIDM)*. IEEE, 2011, pp. 310–317.
- [19] S. K. van den Broucke and J. De Weerd, "Fodina: a robust and flexible heuristic process discovery technique," *decision support systems*, vol. 100, pp. 109–118, 2017.
- [20] J. C. Buijs, B. F. Van Dongen, and W. M. van der Aalst, "On the role of fitness, precision, generalization and simplicity in process discovery," in *OTM Confederated International Conferences "On the Move to Meaningful Internet Systems"*. Springer, 2012, pp. 305–322.
- [21] A. Augusto, R. Conforti, M. Dumas, M. La Rosa, and G. Bruno, "Automated discovery of structured process models: Discover structured vs. discover and structure," in *International Conference on Conceptual Modeling*. Springer, 2016, pp. 313–329.
- [22] W. M. van der Aalst, *Process mining: discovery, conformance and enhancement of business processes*. Springer, 2011, vol. 2.
- [23] A. Augusto, R. Conforti, M. Dumas, and M. La Rosa, "Split miner: Discovering accurate and simple business process models from event logs," in *2017 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2017, pp. 1–10.
- [24] P. Wiśniewski, K. Kluza, and A. Ligeza, "An approach to participatory business process modeling: BPMN model generation using constraint programming and graph composition," *Applied Sciences*, vol. 8, no. 9, p. 1428, 2018.
- [25] M. L. Owoc *et al.*, "Benefits of knowledge acquisition systems for management. an empirical study," in *2015 Federated Conference on Computer Science and Information Systems (FedCSIS)*. IEEE, 2015, pp. 1691–1698.
- [26] W. M. van der Aalst, T. Weijters, and L. Maruster, "Workflow mining: Discovering process models from event logs," *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 9, pp. 1128–1142, 2004.
- [27] E. Tsang, *Foundations of constraint satisfaction: the classic text*. BoD–Books on Demand, 2014.
- [28] A. Niederliński, *Programowanie w logice z ograniczeniami: Łagodne wprowadzenie dla platformy ECLiPSe*. Wydawnictwo Pracowni Komputerowej Jacka Skalmierskiego, 2010.
- [29] P. Wiśniewski, K. Kluza, M. Słazyński, and A. Ligeza, "Constraint-based composition of business process models," in *Business Process Management Workshops*, E. Teniente and M. Weidlich, Eds. Cham: Springer International Publishing, 2018, pp. 133–141.
- [30] A. A. Cervantes, N. R. van Beest, M. La Rosa, M. Dumas, and L. García-Bañuelos, "Interactive and incremental business process model repair," in *OTM Confederated International Conferences "On the Move to Meaningful Internet Systems"*. Springer, 2017, pp. 53–74.
- [31] B. F. Van Dongen, A. K. A. de Medeiros, H. Verbeek, A. Weijters, and W. M. Van Der Aalst, "The ProM framework: A new era in process mining tool support," in *International conference on application and theory of petri nets*. Springer, 2005, pp. 444–454.

# Analysis of the Correlation Between Personal Factors and Visiting Locations With Boosting Technique

Ha Yoon Song

Department of Computer Engineering  
Hongik University, Seoul, Republic of Korea  
Email: hayoon@hongik.ac.kr

JiSeon Yun

Department of Computer Engineering,  
Hongik University, Seoul, Republic of Korea  
Email: wltjs9214@gmail.com

**Abstract**—The paper analyzed the relationship between the person’s fourteen characteristic factors and place to visit. The personal factors consist of personality, marital Status, final education, majors, religion, monthly income, commuting means and time, frequency of travel, userage of social media, time spent on social media per day, cultural type. In addition, the analysis was done on which factors have the greatest impact. The analysis involved thirty-four participants and the boosting technique was used as a method of analysis.

## I. INTRODUCTION

RECENTLY, A number of services provides useful information to people by predicting their moving pattern and location data, especially for Location Based Service (LBS). However, most of the studies predicting people’s movements focus on analyzing past patterns of movement. Apart from this prediction method, we conducted another research on a relationship where a person visits with person’s various factors [1] [2]. Factors such as a personality, marital status, and final education and so on clearly affect a person’s favorite place to visit. In this study, the correlation between person’s characteristic factors and place to visit are analyzed using Boosting techniques. In addition, the analysis of the greatest influential factors to location visit is also addressed. Section II will describe the Boosting technique to be used for correlation analysis. Section III will describe the person’s characteristic data and location data used in the analysis. Section IV analyzes which factors have the greatest impact. Section V will describe the conclusions of this study and the future direction of study.

## II. BOOSTING AS AN ANALYSIS METHOD

### A. Boosting

The analysis technique used for this study is Boosting, one of the ensembles techniques [3]. Boosting is one of the techniques of generating a number of classifiers by manipulating initial sample data similar to Bagging, but the biggest difference is that Boosting is a sequential method. Boosting is a technique to train several weak learners sequentially, to

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST)(NRF-2017R1D1A1B03029788).

TABLE I  
PERSONALITY DATA OF 5 VOLUNTEERS USING BFF

	O	C	E	A	N
Person1	3.3	3.9	3.3	3.7	2.6
Person2	2.7	3.2	3.2	2.7	2.8
Person3	4.3	3.1	2.3	3.2	2.9
Person4	4.2	4.3	3.5	3.6	2.6
Person5	4.0	3.7	4.0	3.9	2.8

learn weight by adding weight to mispredicted data, and to predict using the finally generated learner. That is, the results of the previous learning will affect the next learning.

### B. XGBoost

There are a number of Boosting algorithms. In this research, we will use XGBoost boosting algorithm. XGBoost is an algorithm that visualizes how much the model relies on which factors [4] [5]. It also offers a variety of custom optimization options, including evaluation functions for flexibility. Therefore, it was appropriate to analyze which factors have the greatest impact on place to visit.

## III. INPUT DATA

### A. Personality Data

Personality data was digitized into five personality types in the Big Five Factor (BFF). BFF was developed by psychologists P. T. Costa and R. McCrae in 1976 and is a personality psychological model that explains human personality in terms of five mutually independent factors [6] [7] [8] [9] [10] [11] [12]. O is Openness, C is Conscientiousness, E is Extroversion, A is Agreeableness, and N is Neuroticism. Table I is personality data of 5 volunteers using BFF.

### B. Other Personal Factors

The person’s factor without personality were collected through a questionnaire made directly by Google Form and quantified the categories for each factor. Table II is the person’s characteristic factors without personality of four volunteers

obtained from the questionnaire. Age refers to age, with 1 in the teens, 2 in the 20s, 3 in the 30s and 4 in the 40s and older. Job represents a job and has been assigned a category by adding 'students' to the International Classification of Work (ISCO) standard [13]. 1 is for students, 2 is for managers, 3 is for technical workers, 5 is for office workers, 6 is for service and sales, 7 is for functional workers, 8 is for device and machine operation, and 9 is for simple labor workers. Marriage indicates marital status, 1 is married and 2 is unmarried. Edu represents final education, 1 is below high school graduation, 2 is a high school graduate, 3 is a university graduate, 4 is master's degree and 5 is doctoral degree. Major represents the major, 1 is the humanities, 2 is the sociality, 3 is the educational, 4 is the engineering, 5 is the natural science, 6 is the medicine and 7 is the art. Religion represents religion, 1 is Atheist, 2 is Christianity, 3 is Catholicism (the Catholic Church), and 4 is Buddhism. Salary represents monthly income, with 1 being less than 500,000 won, 2 being less than 1 million won, 3 being more than 1 million won, 4 being more than 2 million won and 5 being more than 3 million won. Vehicle indicates means of commuting, 1 is walking, 2 is cycling, 3 is using self-driving, and 4 is public transportation. Comm T indicates commuting time, 1 is within 30 minutes, 2 is less than one hour, 3 is less than one hour, and 4 is more than two hours. Travel indicates the frequency of travel, 1 is less than 1 time, 2 is less than 4 times, 3 is less than 4 times, and 4 is more than 6 times. Social M indicates usage of social media, 1 is on social media, and 2 is not on social media. Social M2 represents the daily usage of social media, 1 is less than 30 minutes, 2 is less than 1 hour for 30 minutes, 3 is less than 1 hour and 4 is more than 3 hours. Finally, Culture represents cultural type, 1 corresponds to a mixture of static activity, 2 to dynamic activity, and 3 to both static and dynamic activities.

TABLE II  
PERSON'S FEATURE DATA FROM THE QUESTIONNAIRE

	Volunteer1	Volunteer2	Volunteer3	Volunteer4
Age	2	2	3	2
Job	1	1	3	1
Marriage	2	2	2	2
Edu	2	2	4	4
Major	4	4	4	4
Religion	1	3	2	4
Salary	2	2	5	2
Vehicles	4	4	2	4
Comm T	3	3	2	3
Travel	2	2	2	3
Social M	1	2	2	1
Social M2	3	0	0	2
Culture	3	3	2	2

### C. Location Categories

The SWARM application was used to collect location data. SWARM is an application that records the location of a visit when a user visits a site. Location data was created by categorizing each visit data to ten industry classification and accumulating number of visits for each category [14] [15]. Ten Industry categories include Foreign Institutions, Retail, Service industry, etc. Finally, Location data is obtained by calculating the ratio of number of visits of each category compared to the total number of visits. Table III is part of location data of four volunteers.

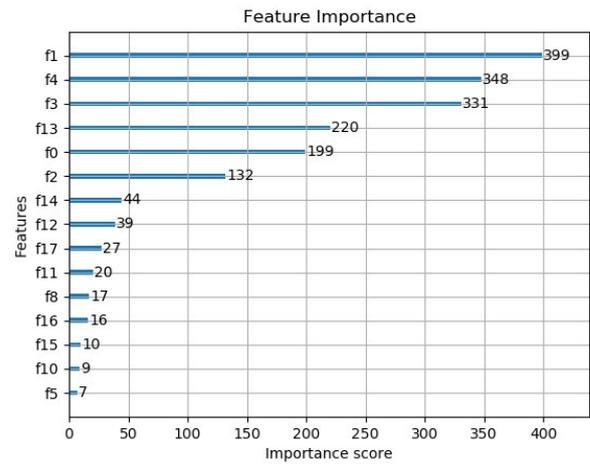


Fig. 1. Feature Importance Graph for Location Categories

## IV. EXPERIMENTAL RESULT

We used XGBoost mentioned in section II as an analysis technique. An independent variable is a person's characteristic data, which was created by merging personality data obtained using BFF, and data for the rest of the factors obtained through a questionnaire. Table IV is characteristic data of three volunteers. Dependent variable is location data. A regression model was created by inserting dependent variable and independent variables into `XGBRegressor()` in XGBoost. Then, `feature_importances` on this regression model and found what factors among the various characteristics of person including personality are most effective for location data.

Figs. 1 show the result of performing feature importance analysis using XGBoost for Foreign Institutions. The y-axis (Features) represents each factor included in the person's characteristic factors. The x-axis (Importance) represents the effectiveness of the independent variable for the dependent variable. Labels f0 through f17 are in the order of the factors listed in Table IV. For example, in Fig. 1, Foreign Institutions of Feature Importance shows that f1 (C, Conscientiousness) has the greatest impact on location data classified as Foreign institutions.

TABLE III  
SAMPLE LOCATION CATEGORY VISITING RATES OF FOUR VOLUNTEERS

	Volunteer1	Volunteer2	Volunteer3	Volunteer4
<b>Foreign Institutions</b>	0.01705	0.00551	0.13559	0.25833
<b>Retail</b>	0.05634	0.67250	0.04237	0.01667
<b>Service industry</b>	0.02965	0.00162	0.02260	0.00333
<b>Restaurant</b>	0.19496	0.07620	0.40960	0.15167
<b>Pub</b>	0.02743	0.01232	0.00847	0.02000
<b>Cafe</b>	0.19422	0.07847	0.07910	0.06167
<b>Cinema</b>	0.01705	0.00551	0.00565	0.01000
<b>Educational institution</b>	0.43662	0.14008	0.27401	0.47333
<b>Hospital</b>	0.00741	0.00292	0.02260	0.00000
<b>Historic sites</b>	0.01927	0.00486	0.00000	0.00500

## V. CONCLUSION

In this study, we analyzed the correlation between various factors of people and place to visit through boosting. As a result, we were able to see how each characteristic of a person affects each place visit. However, there are many similar results for each of the ten place data, and the accuracy was not high. Therefore, we analyzed the reason in many ways. Firstly, many biased results were obtained because most of the volunteer were students in the process of collecting data. Therefore, in the next study, we will recruit the volunteer by various occupations and ages. Secondly, the number of volunteers was few, and the data of place to visit were also insufficient. This is because the volunteer does not use the SWARM application properly. SWARM does not automatically collect the places visited, but it is inconvenient because user has to check-in themselves actively. Therefore, in the next study, we

will recruit more volunteers and make detailed guidance on how to collect data with SWARM. Thirdly, there are several parameters when generating the XGBoost predictive model. When using XGBoost, tuning hyperparameters means that they are the most essential and important. There might be more way to tune XGBoost parameters for future research. Lastly, the accuracy is not great because most ambiguous places to sort are put into service industry or Historic sites in the process of applying the visited places to the industry classification. Therefore, location category classification should be improved by other than current industry classification standards.

Location-based services (LBS) is one of the emerging issues that have great potential for future service. In particular, understanding human mobility patterns is a key part of LBS. We can analyze human mobility patterns by using the correlations between various factors of people and visiting locations analyzed in this study. Therefore, this analysis result might be extended and can be utilized in LBS. It is also expected to be useful for recommendation systems. A recommendation system is a kind of information filtering technology that recommends information that might be of interest to a specific user, such as video recommendations of Netflix and YouTube. People with specific factors will be able to correlate the frequent visits to specific places and apply them to the recommendation system.

Location-based services (LBS) is one of the emerging issues that have great potential for future service. In particular, understanding human mobility patterns is a key part of LBS. We can analyze human mobility patterns by using the correlations between various factors of people and visiting locations analyzed in this study. It is also expected to be useful for recommendation systems. A recommendation system is a kind of information filtering technology that recommends information that might be of interest to a specific user. People with specific factors will be able to correlate the frequent visits to specific places and apply them to the recommendation system.

## REFERENCES

- [1] S. Y. Kim and H. Y. Song, "Predicting human location based on human personality," *International Conference on Next Generation Wired/Wireless Networking*, 2014. doi: <https://doi.org/10.1007/978-3-319-10353-2-7>

TABLE IV  
PERSONAL FACTORS OF THREE VOLUNTEERS

	Volunteer1	Volunteer2	Volunteer3
O (f0)	3.3	2.7	4.3
C (f1)	3.9	3.2	3.1
E (f2)	3.3	3.2	2.3
A (f3)	3.7	2.7	3.2
N (f4)	2.6	2.8	2.9
Age (f5)	2	2	3
Job (f6)	1	1	3
Marriage (f7)	2	2	2
Edu (f8)	2	2	4
Major (f9)	4	4	4
Religion (f10)	1	3	2
Salary (f11)	2	2	5
Vehicles (f12)	4	4	2
Comm T (f13)	3	3	2
Travel (f14)	2	2	2
Social M (f15)	1	2	2
Social M2 (f16)	3	0	0
Culture (f17)	3	3	2

TABLE V  
FEATURE IMPORTANCE OF EACH LOCATION CATEGORIES

	feature 1	feature 2	feature 3	feature 4	feature 5
Foreign Institutions	C	N	A	Comm T	O
	0.219472	0.191419	0.182068	0.121012	0.109461
Retail	O	N	Travel	E	Social M2
	0.243421	0.154605	0.115132	0.108553	0.105263
Service industry	C	O	Social M2	Edu	Comm T
	0.250000	0.180147	0.161765	0.147059	0.088235
Restaurant	O	A	Vehicles	Salary	E
	0.255132	0.184751	0.114370	0.102639	0.099707
Pub	O	Salary	C	N	A
	0.367742	0.258065	0.129032	0.109677	0.064516
Cafe	O	C	N	Salary	E
	0.255435	0.217391	0.125000	0.089674	0.084239
Cinema	C	N	A	Vehicles	Religion
	0.243590	0.153846	0.123932	0.085470	0.085470
Educational institution	O	C	Salary	Edu	N
	0.215827	0.165468	0.158273	0.146283	0.083933
Hospital	O	Salary	C	Travel	A
	0.247664	0.219626	0.140187	0.126168	0.079439
Historic sites	O	Salary	N	C	Comm T
	0.219780	0.179487	0.175824	0.157509	0.131868

- [2] M. J. Chorley, R. M. Whitaker, and S. M. Allen, "Personality and location-based social networks," *Computers in Human Behavior*, vol. 46, pp. 45 – 56, 2015. doi: <https://doi.org/10.1016/j.chb.2014.12.038>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0747563214007559>
- [3] F. Schapire, Robert E, *Boosting: Foundations and Algorithms*. MIT Press (MA), 2014.
- [4] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," *CoRR*, vol. abs/1603.02754, 2016. doi: 10.1145/2939672.2939785. [Online]. Available: <http://arxiv.org/abs/1603.02754>
- [5] "Xgboost," <https://xgboost.readthedocs.io/en/latest/index.html>, accessed: 2019-01-15.
- [6] P. T. Costa and R. R. McCrae, "Four ways five factors are basic," *Personality and Individual Differences*, vol. 13, no. 6, pp. 653 – 665, 1992. doi: [https://doi.org/10.1016/0191-8869\(92\)90236-I](https://doi.org/10.1016/0191-8869(92)90236-I). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/0191886992902361>
- [7] J. Hoseinifar, M. M. Siedkalan, S. R. Zirak, M. Nowrozi, A. Shaker, E. Meamar, and E. Ghaderi, "An investigation of the relation between creativity and five factors of personality in students," *Procedia - Social and Behavioral Sciences*, vol. 30, pp. 2037 – 2041, 2011. doi: <https://doi.org/10.1016/j.sbspro.2011.10.394> 2nd World Conference on Psychology, Counselling and Guidance - 2011. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1877042811022191>
- [8] D. Jani, J.-H. Jang, and Y.-H. Hwang, "Big five factors of personality and tourists' internet search behavior," *Asia Pacific Journal of Tourism Research*, vol. 19, 05 2014. doi: 10.1080/10941665.2013.773922
- [9] D. Jani and H. Han, "Personality, social comparison, consumption emotions, satisfaction, and behavioral intentions: How do these and other factors relate in a hotel setting?" *International Journal of Contemporary Hospitality Management*, vol. 25, no. 7, pp. 970–993, 2013. doi: 10.1108/IJCHM-10-2012-0183. [Online]. Available: <https://doi.org/10.1108/IJCHM-10-2012-0183>
- [10] O. P. John and S. Srivastava, "The big-five trait taxonomy: History, measurement, and theoretical perspectives," 1999.
- [11] L. R. Goldberg, "'the structure of phenotypic personality traits": Author's reactions to the six comments." *American Psychologist*, vol. 48, pp. 1303–1304, 12 1993. doi: 10.1037/0003-066X.48.12.1303
- [12] Y. Amichai-Hamburger and G. Vinitzky, "Social network use and personality," *Computers in Human Behavior*, vol. 26, no. 6, pp. 1289 – 1295, 2010. doi: <https://doi.org/10.1016/j.chb.2010.03.018> Online Inter-activity: Role of Technology in Behavior Change. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0747563210000580>
- [13] "International standard classification of occupation (isco)," <https://www.ilo.org/>, accessed: 2018-12-20.
- [14] E. B. Lee and H. Y. Song, "An analysis of the relationship between human personality and favored location," *The Seventh International Conference on Advances in Future Internet*, 2015.
- [15] Song, Ha Yoon and Kang, Hwa Baek, "Analysis of relationship between personality and favorite places with poisson regression analysis," *ITM Web Conf.*, vol. 16, p. 02001, 2018. doi: 10.1051/itmconf/20181602001. [Online]. Available: <https://doi.org/10.1051/itmconf/20181602001>

# Do online reviews reveal mobile application usability and user experience? The case of WhatsApp

Paweł Weichbroth  
WSB University in Gdansk,  
Grunwaldzka 238A,  
82-266 Gdansk, Poland  
Email: pweichbroth@wsb.gda.pl

Anna Baj-Rogowska  
Gdansk University of Technology,  
Narutowicza 11/12,  
80-233 Gdansk, Poland  
Email: anna.baj-rogowska@zie.pg.gda.pl<sup>✉</sup>

**Abstract**—The variety of hardware devices and the diversity of their users imposes new requirements and expectations on designers and developers of mobile applications (apps). While the Internet has enabled new forms of communication platform, online stores provide the ability to review apps. These informal online app reviews have become a viral form of electronic word-of-mouth (eWOM), covering a plethora of issues. In our study, we set ourselves the goal of investigating whether online reviews reveal usability and user experience (UUX) issues, being important quality-in-use characteristics. To address this problem, we used sentiment analysis techniques, with the aim of extracting relevant keywords from eWOM WhatsApp data. Based on the extracted keywords, we next identified the original users' reviews, and individually assigned each attribute and dimension to them. Eventually, the reported issues were thematically synthesized into 7 attributes and 8 dimensions. If one asks whether online reviews reveal genuine UUX issues, in this case, the answer is definitely affirmative.

## I. INTRODUCTION

WITH the rapid development of mobile devices, increasing numbers of mobile applications (apps) are being manufactured and deployed, and these apps are accompanied by rich user reviews. This informal type of communication, directed at an unspecified number of people using internet-based technology and related to the usage of particular goods or services is defined as electronic word-of-mouth (eWOM) [1]. Undeniably, this phenomenon has attracted considerable attention from application users as well as their vendors. According to Mobile App Daily [2], the most trusted and largest media source of the mobile app industry, more than 70 percent of people read app reviews before downloading, while, more importantly, 75 percent identified reviews as a key driver for downloading, and 42 percent consider app store reviews as equally or more trustworthy than personal recommendations [3].

Inspired by these findings, in our study we investigate the content of online reviews. The broad area of topics gaudily reported by users roughly corresponds to a similar number of application properties. Therefore, in this study the focus is on quality-in-use issues, which are recognized as the subject of interest of usability and user experience (UUX) practitioners. The evaluation of the UUX of a mobile application has been identified by many as one of the main challenges [4,5,6], eventually determining the success of its continued acceptance by users.

On the other hand, while the majority of recent studies on the perceived quality of an app have focused on quality assurance from the perspective of its development or testing, this study, on the contrary, solely concentrates on the end user's attitude to an app, expressed by eWOM. In particular, we put forward one research question: do online reviews reveal mobile application usability and user experience? In other words, by assumption, we attempted to extract valuable information from eWOM data concerning the facets of UUX.

The remainder of this paper is structured as follows. We first review the background and relevant literature in Section 2. Sections 3 and 4 introduce the research methodology and experimental setup, respectively. Section 5 presents the empirical results obtained in the study, followed by a discussion of the findings and implications, given in Section 6. Finally, Section 7 concludes the study.

## II. THEORETICAL BACKGROUND AND RELATED WORK

In the light of the results obtained in our previous study [7], in the context of mobile applications, the majority of studies have pointed to the usability definition adapted from the ISO 9241-11 norm. Here, usability is defined as “the extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” [8]. Furthermore, along with these three already articulated attributes, in some studies, other attributes have also been considered, namely: learnability, memorability, cognitive load, errors, ease of use, navigation and operability [7].

Under the umbrella of user experience, all of a “person's perceptions and responses resulting from the use and/or anticipated use of a product, system or service” [9] are a subject of concern. Based on the existing body of knowledge [10], we elaborated a list of UX dimensions, from which we elected eight unique dimensions: aesthetics, enjoyment, hedonics, trust, support, engagement, discomfort and frustration.

It is worth noting that, according to the above norm, usability, when interpreted from the perspective of a user's personal goals, can include the kind of perceptual and emotional aspects typically associated with user experience. Moreover, usability criteria can also be used to evaluate aspects of user experience.

To capture usability and/or user experience, there are two not mutually exclusive approaches [11], which are applicable either during or after application usage. The former mainly concerns laboratory testing, while the latter is a retrospective analysis of data, gathered in the form of a questionnaire [12,13,14], video recording [15,16,17], or, more notably, online reviews [18,19,20,21].

Jacob and Harrison [18] argue that 23.3 percent of mobile app reviews represent feature requests, where users either suggested new features or expressed their preferences for the re-design of existing ones. The prototype experimental tool (MARA) was used to mine and retrieve feature requests from the data of online reviews. In particular, the data are processed in a fixed sequence: review retrieval, feature request mining, feature request summarization, and feature request visualization. During the first phase, a web crawler extracts the source page which contains the reviews of a given app and parses their content. The meta-data, including the posting date, the user's rating, and other fields, are also collected. The meta-data, as well as the content of the review are normalized to reduce noise in the final results, where the latter is also split into sentences, using [22], a toolkit for processing text by use of computational linguistics. The second phase uses the split review content as input and mines for feature requests expressed by users. The mining algorithm utilizes a set of linguistic rules defined for supporting the identification of sentences which refer to particular requests. During the third phase, the system summarizes the extracted feature requests according to a set of predefined rules. The applied rules aim to rank the extracted user requests based on their frequency and length. The more frequent and lengthier feature requests would be first in the summary. Finally, during the visualization phase, the results of the summarization are displayed to the user.

He et al. [19] propose a feature-opinion mining approach to automatically summarize the reviews, based on dependency parsing. The approach utilizes a regression model to generate sentiment words, consisting of a phrase and its sentiment weight. Next, the feature is extracted, based on the dependency relationship between the feature and sentiment words. Eventually, a score is assigned to the feature according to the dependency relationship. In general, the applied approach consists of three phases: (1) sentiment word generation, (2) feature extraction, and (3) feature scoring.

Jin et al. [20] illustrate a framework to select pairs of opinionated representative yet comparative sentences with specific product features from online reviews of competitive products. Sentiment analysis techniques were applied to identify opinionated sentences referring to a specific feature from product online reviews. To select a "small" number of representative yet comparative opinionated sentences from those identified, the authors investigated the representativeness, comparativeness and diversity of the information. The contribution of this study lies in three

greedy algorithms to analyse the optimization problem for suboptimal solutions.

A comprehensive study of existing solutions for mining online opinions is given by [21]. There are several methods identified, including LDA (Latent Dirichlet Allocation), ASUM (Aspect and Sentiment Unification model), statistical analysis, SVM (Support Vector Machine), EMNB (Expectation Maximization for Naïve Bayes), decision trees, manual tagging, keyword extraction with grouping and ranking, and others.

To sum up, having briefly depicted the main ideas from arbitrarily selected studies, in this study we performed a semi-automated review analysis, methodologically similar to the framework developed by Vu et al. [23].

### III. METHODOLOGY

In our study, the sentiment analysis is aided by the WordStat Sentiment Dictionary, designed by combining negative and positive words from three different sources: Harvard IV dictionary, the Regressive Imagery Dictionary (RID) and the Linguistic and Word Count dictionary. Eventually, more than 9526 negative and 4669 positive word patterns were gathered [24].

A user's sentiment is not measured by those two lists of words and word patterns, but instead by two sets of rules which are intended to take into account the negations preceding those words. For example, negative sentiment is measured by applying the following two rules:

- negative words are not preceded by a negation (e.g. no, not, never) within four words in the same sentence;
- positive words are preceded by a negation within four words in the same sentence.

On the other hand, positive sentiment is measured in a similar way by alternatively checking the following two rules:

- positive words are not preceded by a negation;
- negative terms are followed by a negation.

However, some argue that the latter rule shows less predictive properties, and in some cases, might even deteriorate the sentiment measurement [25,26].

In general, the sentiment analysis was carried out in a fixed sequence of five stages [27], as depicted and described below (Fig. 1):



Fig. 1 The sentiment analysis process

Data collection (A) involves downloading the text data from the Web and assembling one consolidated data set.

Text preparation (B) aims to clean and transform the collected data, comprising the following two tasks:

- data parsing, which means analyzing data and breaking them down into smaller blocks, which separately can be easily interpreted and managed, and

- data pre-processing, which concerns: (i) performing tokenization, where the words are transformed from the text into a structured set of elements (tokens); (ii) executing a stop word list, where the words which have low informative value or are semantically insignificant (e.g. *and*, *also*, *or*) are removed; and (iii) reducing the words by individually extracting a stem word (a root of words).

Sentiment detection (C) is to identify sentences with subjective expressions (opinions, beliefs and views) and to reject objective communication sentences (facts, factual information).

Sentiment classification (D) is the task of classifying a text in a document into a positive or negative class on various levels (e.g. document, sentence and aspect of entities).

Visualization of output (E) aims at transforming data, information and knowledge into a visual form (e.g. pie, bar, line graph) to take advantage of natural human visual capabilities [28,29,30].

In the next step, we assumed, after [31], that in textual analysis research, a higher negative (positive) word frequency indicates a more pessimistic (optimistic) sentiment. Therefore, we extracted all negative and positive words with the highest frequency of occurrence. Next, we consequently mapped these words to a particular usability attribute and/or user experience dimension. Finally, identifying the original reviews, based on keyword searching, enabled us to individually assign them to the relevant attributes and dimensions.

#### IV. EXPERIMENTAL SETUP

In total, we collected 399 reviews by WhatsApp users from the Google Play website using a self-made web crawler. The data set is both human and computer readable due to the JSON (JavaScript Object Notation) format applied.

Let  $i = \{1, 2, \dots, n\}$  be the ordinal number of a user's review. Each review can be defined as a set of six variables (s sextuple):

$review(i) = \{name: string, rate: [1-5], when: date, helpful: integer, short-review: string, full-review: string\}$ , where:

- *name* is the name of a user (reviewer), which may consist of first name, surname or any other string of characters (e.g. John, John Kowalski, JK);
- *rate* is the numerical evaluation of the mobile application in the range of 1 to 5, given by a user,
- *when* is the date of the rate, written in a short format (e.g. February 12, 2019),
- *helpful* is the number of thumbs-ups given by users for the review,
- *short-review* is a verbal evaluation of the mobile app,
- *full-review* is also a verbal evaluation of the mobile application, with a higher number of characters allowed.

It is worth noting that a user can add a review to a particular app if it has been downloaded and installed.

The sentiment analysis was conducted using the ProSuite commercial software [32], being an integrated collection of Provalis Research Text Analytics Tools that allow one to explore, analyse and relate both structured and unstructured data. The computing platform includes three major tools:

- QDA Miner for qualitative data analysis, including coding, annotating, retrieving and analyzing small and large collections of documents and images;
- WordStat for the content analysis of open-ended responses, interview or focus group transcripts, for information extraction and knowledge discovery from incident reports and customer complaints, and for the automatic tagging and classification of documents;
- SimStat for statistical analysis, supporting both numerical and categorical data, dates and short alpha-numeric variables, as well as memo and document variables.

These tools have also been used in other studies for content analysis and text mining [33,34,35], allowing researchers to integrate numerical and textual data into a single project.

#### V. RESULTS

The research material constituting reviews by WhatsApp mobile application users created a so-called bag-of-words (BOW). After transforming the text into a BOW, we can calculate various measures to characterize the text. In our study, the BOW model consists of 4 245 words (tokens). The most common words are shown below (Fig. 2).

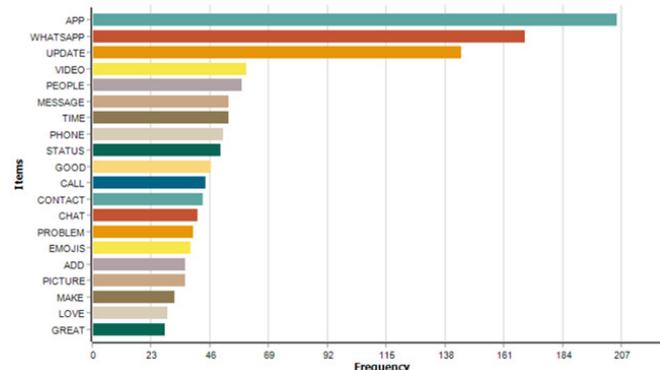


Fig. 2 The distribution of keywords by frequency

In the first step, the sentiment analysis was performed on the users' opinions. The sentiment analysis, conducted according to the stages shown in Figure 1, contained 904 negative words (21.30%) and 1217 positive items (28.67%). However, neutral words identified in the study (50.03%) can be ignored because they do not add value to the study. The obtained results are given below (Fig. 3).

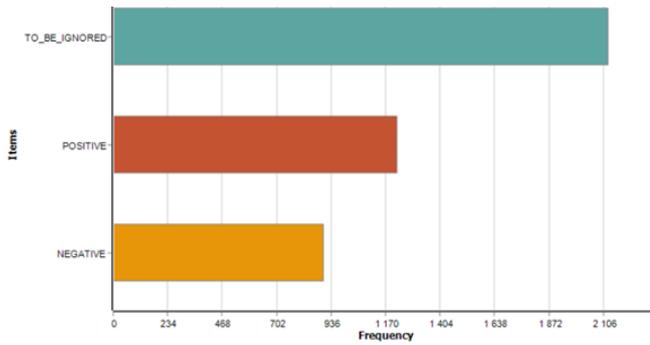


Fig. 3 The distribution of words after the process of sentiment analysis

In the WhatsApp users' ratings, the advantage of positive sentiment is clearly visible. The set of bigrams extracted from the users' reviews also show this trend (e.g. *great app*, *excellent app*, *good app*, *love WhatsApp*).

In the next step, we assumed, as already indicated above, that in textual analysis research, a higher negative (positive) word frequency indicates a more pessimistic (optimistic) sentiment. Therefore, we extracted the crucial negative and positive words with the highest frequency of occurrence (Table 1).

TABLE I.  
LIST OF THE MOST FREQUENT KEYWORDS

Negative		Positive	
Word	Frequency	Word	Frequency
fix	46	call	58
problem	39	contact	57
issue	36	good	46
number	19	friend	34
annoying	18	make	32
remove	18	feature	32
bug	17	share	31
hate	13	love	29
unable	11	work	29
stop	11	great	28
bad	11	open	15
delete	10	fine	14
sucks	9	easy	12
horrible	9	quality	12
reduce	8	make	11
lost	8	nice	11
error	8	free	10
limit	7	happy	9
wrong	6	awesome	9
miss	6	excellent	8

Keywords frequently appearing with negative reviews are likely to describe the issues or features of apps that cause a negative user experience, i.e. making users unsatisfied. Thus, such keywords would be of interest to app developers because they can help to identify the bad aspects of an app

and user opinions about such aspects (e.g. bigrams: *fix this issue*, *Feb update*, *app lock*, *dark mode*). And similarly, with positive words that point to the aspects that satisfy the user (e.g. bigrams: *good work*, *excellent app*, *good app*).

Based on selected keywords from the sentiment analysis, we searched for the actual user reviews that are the most relevant to those keywords. On this basis, we were able to assign usability attributes and user experience dimensions. An example of our work is included in Table 5 and Table 6 (see Appendix). The mappings between keywords and usability (Fig. 4) and UX dimensions (Fig. 5) are shown below.

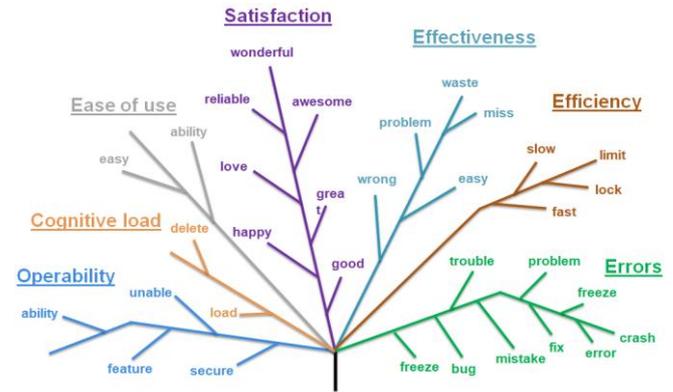


Fig. 4. The mapping between keywords and usability attributes

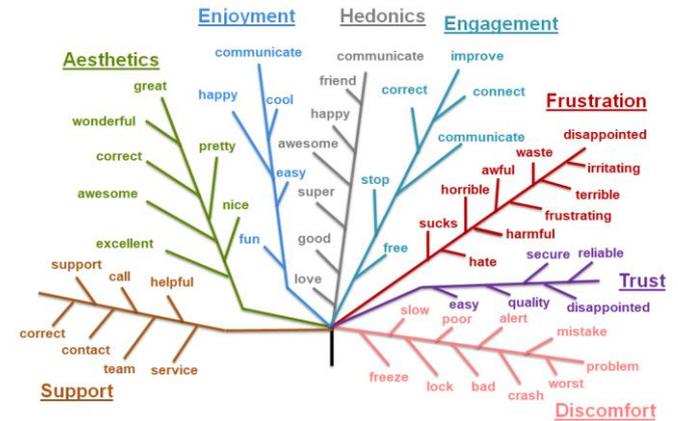


Fig. 5. The mapping between keywords and UX dimensions

The same keywords (e.g. *easy*, *awesome*, *problem* and *communicate*) can be used to describe different UUX attributes and dimensions. In addition, the negative words describe a larger number of dimensions and attributes than the positive ones. Similar conclusions were drawn by Provost and Robert [36].

Next, the bag of words was divided into seven clusters by applying the hierarchical grouping method (Table 2), where a cluster is a group (or class) of similar objects created as a result of data grouping. On further analysis, these clusters can be compared to the classes developed by grouping original user reviews, which have an absolute meaning and should not be standardized.

TABLE II.  
KEYWORD CLUSTERS

No	Keywords
1	problem, contact, message, send, people, time, fix, good, profile, chat, option, video, phone
2	unable, user, text, notification, nice, long, full, check, call, bug, communication, feature, easy, support, data, bad, delete, issue, quality, post, friend
3	view, team, set, screen, message, card, conversation
4	work, update, online, chat, fine, hate, voice, annoying, video, app, feature, photo
5	call, thing, change, friend, person, issue, great, version, group, picture, love
6	download, free
7	update, WhatsApp

The steps completed so far have provided a basis for the mapping of frequent keywords to usability attributes and user experience dimensions (Table 3 and Table 4).

TABLE III.  
THE MAPPING BETWEEN FREQUENT KEYWORDS AND USABILITY ATTRIBUTES

Attribute	Keywords
efficiency	limit, slow, lock, fast
satisfaction	good, love, great, awesome, happy, reliable, wonderful
effectiveness	wrong, miss, problem, easy, waste
learnability	–
memorability	–
cognitive load	delete, load
errors	bug, error, fix, freeze, crash, problem, trouble, mistake
ease of use	ability, easy
operability	ability, unable, secure, feature

TABLE IV.  
THE MAPPING BETWEEN FREQUENT KEYWORDS AND UX DIMENSIONS

Dimension	Keywords
aesthetics	correct, great, nice, awesome, excellent, pretty, wonderful
enjoyment	happy, fun, cool, easy, communicate
hedonics	friend, communicate, awesome, love, happy, super, good,
trust	quality, secure, reliable, easy
support	service, contact, call, support, team, helpful, correct
engagement	correct, stop, communicate, connect, improve, free
discomfort	alert, problem, bad, slow, poor, lock, freeze, crash, worst, mistake, disappointed
frustration	irritating, sucks, horrible, waste, disappointed, terrible, hate, irritating, frustrating, awful, harmful

Interestingly, two usability attributes are empty sets. In other words, none of the keywords were assigned, which indicates that users neither report on the ability to learn nor to remember. Moreover, one can classify UX dimensions as positive (aesthetics, enjoyment, hedonics, trust, support), neutral (engagement) and negative (discomfort, frustration). In Fig. 5 specific dimensions were marked off by labelling sets of the keywords with different colors, ranging from green and blue to red, respectively.

On the other hand, the words included in the above two tables indicate the importance of the reported UUX issues by the users. As a matter of fact, eWOM data are meaningful for app vendors not only because users often rely on this resource when making decisions, but more importantly, online reviews might leverage app design and quality.

## VI. DISCUSSION

There is no doubt that the ability provided to users to tell their stories about mobile applications in any way, has brought popularity as well as obstacles for apps. However, there are many examples of those who have taken an unfair advantage this ability. For example, in December 2018, as a response, Google announced a crackdown on app developers who buy ratings and reviews to deceive users or ruin their competitors' reputations [37].

Moreover, in the Notes section of the store, one can read that reviews are automatically processed to find inappropriate content (such as obscene, offensive, or meaningless language). Online reviews are also automatically scanned for spam (like messages sent by bots or repeated content posted multiple times or from multiple accounts). The company has no tolerance for fake reviews, which will be taken down if they are flagged as fake or are in violation of review policies. Therefore, in our opinion, the Google online store of mobile applications is a reliable source of information.

Although 50.03% of identified words were discovered to be valueless, we found the other half of great value. Indeed, eWOM involves positive and negative statements made by users about WhatsApp. This real user-created information has brought insight into users' direct experiences as well as application performance and properties. On the other hand, since software testers are not able to detect all bugs, defects and errors, the users act on their behalf unintentionally but competently.

Like any other similar research, this study has both its limitations as well as strengths. Firstly, only one app, as the source of the reviews, was explored in order to gather the necessary evidence to formulate an answer to the research question. Secondly, there is no mechanism implemented which could automatically process a relatively large volume of data, and set up keyword clusters in a non-supervised mode. Future research will address broadening the sample and implementing a relevant method. Additionally, while the present approach assumed off-line processing, online

processing will also be considered. Lastly, multiple experiments with different apps are being investigated and validated in order to elaborate the unified UX model.

## VII. CONCLUSIONS

In the case of WhatsApp, in this paper we were able to evidence a positive answer to the given research question. eWOM provides a new venue for software vendors to reach users and to influence their opinions. With zero cost for accessing and exchanging information, eWOM creates a new opportunity to better understand users' genuine concerns formulated toward the features and properties of apps, covered by UX theory and practice. In light of the evidenced results, it seems likely that users in increasingly larger numbers will either read and/or write reviews, expecting afterwards to have a better app in the next release.

## REFERENCES

- [1] Ono, A., & Kikumori, M. (2018). Consumer Adoption of PC-Based/Mobile-Based Electronic Word-of-Mouth. In *Encyclopedia of Information Science and Technology*, Fourth Edition, 6019–6030.
- [2] Mobile App Daily (2019). Why Your App Needs Reviews - Importance And Benefits. <https://www.mobileappdaily.com/2018/11/20/why-your-app-need-reviews>
- [3] Walz, A., Ganguly, R. (2015). The Mobile Marketer's Guide to App Store Ratings and Reviews. Apptentive. [http://cdn2.hubspot.net/hubfs/232559/The\\_Mobile\\_Marketers\\_Guide\\_To\\_App\\_Store\\_Ratings\\_and\\_Reviews.pdf](http://cdn2.hubspot.net/hubfs/232559/The_Mobile_Marketers_Guide_To_App_Store_Ratings_and_Reviews.pdf)
- [4] Kukulska-Hulme, A. (2007). Mobile usability and user experience. In *Mobile Learning*, 61–72. Routledge.
- [5] Park, J., Han, S. H., Kim, H. K., Cho, Y., Park, W. (2013). Developing elements of user experience for mobile phones and services: survey, interview, and observation approaches. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 23(4), 279–293.
- [6] Parsazadeh, N., Ali, R., Rezaei, M., Tehrani, S. Z. (2018). The construction and validation of a usability evaluation survey for mobile learning environments. *Studies in Educational Evaluation*, 58, 97–111.
- [7] Weichbroth, P. (2019). Usability of mobile applications: a systematic literature study. (In Print).
- [8] ISO 9241-11 (2018). Ergonomics of human-system interaction – Part 11: Usability: Definitions and concepts.
- [9] ISO 9241-210 (2010), [https://www.iso.org/obp/ui/#iso:std:iso:9241-210:ed-1:v1:en, \[02-04-2019\]](https://www.iso.org/obp/ui/#iso:std:iso:9241-210:ed-1:v1:en, [02-04-2019]).
- [10] Hedegaard, S., Simonsen, J. G. (2013). Extracting usability and user experience information from online user reviews". The SIGCHI Conference on Human Factors in Computing Systems, 2089–2098.
- [11] Korhonen, H., Arrasvuori, J., & Väänänen-Vainio-Mattila, K. (2010). Let users tell the story: evaluating user experience with experience reports. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*, 4051–4056. ACM.
- [12] Lewis, J. R. (1995). IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, 7(1), 57–78.
- [13] Basińska, B., Dąbrowski, D., & Sikorski, M. (2013). Usability and relational factors in user-perceived quality of online services. *Studia Ekonomiczne*, 158, 18–28.
- [14] Moumane, K., Idri, A., & Abran, A. (2016). Usability evaluation of mobile applications using ISO 9241 and ISO 25062 standards. *SpringerPlus*, 5(1), 548.
- [15] Kaikkonen, A., Kekäläinen, A., Cankar, M., Kallio, T., & Kankainen, A. (2005). Usability testing of mobile applications: A comparison between laboratory and field testing. *Journal of Usability studies*, 1(1), 4–16.
- [16] McMillan, D., McGregor, M., & Brown, B. (2015). From in the wild to in vivo: Video Analysis of Mobile Device Use. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, 494–503. ACM.
- [17] Hussain, A., Mkpojiogu, E. O., Jamaludin, N. H., & Moh, S. T. (2017). A usability evaluation of Lazada mobile application. In *AIP Conference Proceedings*, Vol. 1891(1), AIP Publishing.
- [18] Iacob, C., Harrison, R. (2013). Retrieving and analyzing mobile apps feature requests from online reviews. In *Proceedings of the 10th Working Conference on Mining Software Repositories*, 41–44. IEEE.
- [19] He, T., Hao, R., Qi, H., Liu, J., & Wu, Q. (2016). Mining Feature-Opinion from Reviews Based on Dependency Parsing. *International Journal of Software Engineering and Knowledge Engineering*, 26(09n10), 1581–1591.
- [20] Jin, J., Ji, P., & Gu, R. (2016). Identifying comparative customer requirements from product online reviews for competitor analysis. *Engineering Applications of Artificial Intelligence*, 49, 61–73.
- [21] Genc-Nayebi, N., Abran, A. (2017). A systematic literature review: Opinion mining studies from mobile app store user reviews. *Journal of Systems and Software*, 125, 207–219.
- [22] LangPipe. (2019). Home. <http://alias-i.com/lingpipe/index.html>
- [23] Vu, P. M., Nguyen, T. T., Pham, H. V., & Nguyen, T. T. (2015). Mining user opinions in mobile app reviews: A keyword-based approach (t). In *2015 30th IEEE/ACM International Conference on Automated Software Engineering (ASE)*, 749–759. IEEE.
- [24] Provalis Research Group (2019), Sentiment dictionaries, <https://provalisresearch.com/products/content-analysis-software/wordstat-dictionary/sentiment-dictionaries/>, [04-04-2019].
- [25] Wiegand M., Balahur A., Montoyo A., Roth B., Klakow, D. (2010). A Survey on the Role of Negation in Sentiment Analysis, *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, Uppsala, 60–68.
- [26] Asmi, A., Ishaya, T. (2012). Negation Identification and Calculation in Sentiment Analysis. The Second International Conference on Advances in Information Mining and Management.
- [27] Baj-Rogowska A. (2017). Sentiment Analysis of Facebook Posts: the Uber case. *IEEE Eighth International Conference on Intelligent Computing and Information Systems*, pp. 391–395, Cairo (Egypt).
- [28] Weichbroth, P. (2011). The visualisation of association rules in market basket analysis as a supporting method in customer relationship management systems. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, (232), 136–145.
- [29] Pondel, J., & Pondel, M. (2015). BI and Big Data solutions in project management. *Informatyka Ekonomiczna*, 4 (38), 55–63.
- [30] Owoc, M., & Pondel, M. (2016). Selection of Free Software Useful in Business Intelligence. *Teaching Methodology Perspective*. In *IFIP International Workshop on Artificial Intelligence for Knowledge Management*, 93–105. Springer, Cham.
- [31] Loughran T., McDonald B. (2011). When is a liability not a liability? Textual Analysis, Dictionaries and 10-Ks. *The Journal of Finance*, 66(1), pp. 35–66.
- [32] Provalis Research Group (2019), ProSuite, <https://provalisresearch.com/products/qualitative-data-analysis-software/>, [02-03-2019].
- [33] Lock, I., Seele, P. (2016). The credibility of CSR (corporate social responsibility) reports in Europe. Evidence from a quantitative content analysis in 11 countries". *Journal of Cleaner Production*, 122, 186–200.
- [34] Du Plessis, C. (2017). The role of content marketing in social media content communities. *South African Journal of Information Management*, 19(1), 1–7.
- [35] Jones-Diette, J. S., Dean, R. S., Cobb, M., Brennan, M. L. (2019). Validation of text-mining and content analysis techniques using data collected from veterinary practice management software systems in the UK. *Preventive Veterinary Medicine*.
- [36] Provost G., Robert JM. (2013) The Dimensions of Positive and Negative User Experiences with Interactive Products. In: Marcus A. (eds) *Design, User Experience, and Usability. Design Philosophy, Methods, and Tools. DUXU 2013. Lecture Notes in Computer Science*, vol 8012. Springer, Berlin, Heidelberg.
- [37] Cimpanu, C. (2018). Google announces crackdown on Play Store ratings and reviews. <https://www.zdnet.com/article/google-announces-crackdown-on-play-store-ratings-and-reviews/>

## APPENDIX 1

TABLE V.  
EXTRACTED NEGATIVE UUX ATTRIBUTES AND DIMENSIONS FROM ONLINE USERS' REVIEWS

Word	Review	attribute/ dimension
fix	"I am not even getting any notifications from Whatsapp on the status bar nowadays and am very disappointed to say that even though I've double checked the settings for both message and group notifications, there's still no changes. Please <b>fix</b> this problem ASAP."	errors
	"I can't see any status updates from my contacts. The status feature just stops for a while and then returns and stops again. Can you please <b>fix</b> this bug."	errors
	"Useful little app but does come with frustrations. I want to view images and when I click on an image to load it downloads onto my phone which is annoying. Want to view the image not save it and clog up my phone. Same goes for gifs and videos. Please <b>fix</b> it, if you do I'd probably use this more than fb messenger."	errors
	" <b>fix</b> the change! I have contact photos in my phone and people that do not have profile photos would show up with the contact photo now it does not do that after this new update. change it back it was better before"	errors
	"I really like whatsapp messenger but one thing that annoys me is that I cannot forward message to more than (limit set by Whatsapp) five people i guess. Please <b>fix</b> this."	errors
	"do like this app, but the recent update keeps causing it to freeze and crash. Effecting my whole phone. Please <b>fix</b> bugs or whatever is causing it to freeze."	errors
	"Why are whatsapp emojis are looking soooooo badd. like after installing new update, emojis got worse, please <b>fix</b> this in next update."	satisfaction
	"Great App, but there is a bug I am not able to call or video call 3 people at once from the group chat video call option. The call don't respond and automatically disconnect without ringing. Please fix this issue and one more thing when we are getting feature for group video call more then 4 members..."	errors
	"Please <b>Fix</b> bugs . when I Video call , I can't touch anything and can't turn back to the Conversation and can't typing anything . my friend told me either ... please <b>FIX</b> the bugs soon"	errors
problem	"A great way to communicate, but since the last update, my contact's pictures aren't showing. Even though when I go into edit contact and there's a picture there, it's not showing on the main screen. Please <b>fix</b> this!"	errors
	"Plz do something the app has become slow on the two devices I own, one is the huawei p20 and the other is oneplus 6t.I checked my devices but others are all facing the same <b>problem</b> ."	efficiency
	"My WhatsApp crash twice in less than 2 months' time ... All my chats are gone. <b>Problem</b> is I didn't do anything. An error message just pop up and say there's something wrong with my chats history. I lost all my important work chats. This is bad. You can't expect me to do backup every single day."	errors
	"I have a <b>problem</b> sending videos to my contacts, each time I try to send videos that are five minutes long, it is reduced to a lesser minute of 3 minutes of streaming before it can be sent to my contacts. please how do I go about this?"	efficiency
	"I'm using WhatsApp, but I don't see blue coloured double tick after my messages are read. And my friend didn't change the setting on mobile. This not the first time of <b>problem</b> ."	satisfaction
	"Last three weeks I have a <b>problem</b> for message sending. I didn't send any message more than five people. Before 20 people but now 5. I don't know why whatsapp management reduce the conctects for sending messages."	effectiveness
bug	"I can't sent Voice Messages .... Same problem with both my Whatsapp accounts."	errors
	"Everything was fine but since last month's my old chat messages are being deleted by WhatsApp without my knowledge and I am shocked with this new <b>bug</b> ."	errors
	"I can't see any status updates from my contacts.. The status feature just stops for a while and then returns and stops again. Can you please <b>fix</b> this <b>bug</b> ."	satisfaction
hate	"After make video call, it's not getting minimized. Its hanged. New update killing it badly. My mobile note 5 pro.. please solve this <b>bug</b> ."	errors
	"I <b>hate</b> the new update. I lost contact photos to over half my contacts. I can't figure out how to restore contact pics. Also....as popular as this app has been throughout the years, you'd figure that they'd come up with different themes. Instead, same old boring green theme."	satisfaction
	"I <b>hate</b> the new update. I lost contact photos to over half my contacts. I can't figure out how to restore contact pics."	satisfaction
	"I <b>hate</b> the new version. The emojis are old , some are nice , but I would like if the emojis looked more realistic and not fake or something. I recommend this app, although the emojis are not cool. But I would and its useful."	satisfaction
	"I <b>HATE</b> THE NEW UPDATE. ... The previous version was way better. Please restore it."	satisfaction

	"I really love this app...however I and all of my friends <b>hate</b> the new emojis for android... They are awful. Please change the emojis so that they look like IOS emojis ... please"	aesthetics
	"Horrible update. <b>Hate</b> when the update makes the app worse, not better!"	frustration
	"I <b>hate</b> the new update. Pls get back the previous version."	frustration
	"It's still great for communication but I <b>hate</b> the new "upgraded" emojis. As if it wasn't enough that you ruined my favourite emojis, the moons, you've ruined the rolling eyes emoji for me as well. Please fix them and make them look like their past selves."	frustration
<b>bad</b>	"Recent whatsapp update is so <b>bad</b> , I only use it because I have contacts on it. All new icons have gone. Profile icons picked up from phone contacts for those who haven't loaded profile pics, is gone, so there are gaping holes where there should be a contact icon."	discomfort
	"New update is very <b>bad</b> . You can't send msg to more than 5 people, please give us new update and solve this."	satisfaction
	"When sending a video from your gallery, it comes up with an <b>error</b> message ... fix this too."	errors
<b>error</b>	"Unable to send pdf files. <b>Error</b> shows it's not a document what the hell is this.. I think I need to switch messenger."	satisfaction
	"Latest update has a few <b>errors</b> but the one that's bugging me is I had pictures assigned to my phone contacts that used to show as the profile picture on WhatsApp if the person didn't have a profile picture and after the update it's not showing."	errors

TABLE VI.  
EXTRACTED POSITIVE UUX ATTRIBUTES AND DIMENSIONS FROM ONLINE USERS' REVIEWS

<b>good</b>	"My experience is too <b>good</b> with whatsapp. I am happy to make a group and chat within it. It is very helpful and good for school work ... thank you so much."	satisfaction
	"Pros: Its free. Clarity pretty <b>good</b> . Not many adds."	satisfaction
	"It's just so <b>good</b> we can call free, video - chat, share safely, it's one of the necessities in life now. I'm impressed."	satisfaction
<b>feature</b>	"Great! It still remains the most used app in the world. But a <b>feature</b> that can allow us to save what we want needs to be added please."	operate
	"Getting better with each update. The swiping right to reply <b>feature</b> is something I really like."	pleasure
	"... indeed your new <b>features</b> are just amazing. Keep up the good work."	enchantment
	"Neat customization tools, group chat <b>features</b> , and easy location, and now money transfer payments sending & receiving money adding are all cool additions."	comfort
<b>easy</b>	"Fast (especially for sending images), more reliable than SMS, and everyone has it, so it's <b>easy</b> to connect with people."	enjoyment
	"awesome app in social world <b>easy</b> to use fantastic"	easy to use
	"This is the best messaging app ever! I love how it is laid out and how <b>easy</b> it us to use."	easy to use
	"It's quite simply, brilliant. User interface is a tad lame and boring but the app is efficient and <b>easy</b> to use."	efficiency, easy to use
	"Very <b>easy</b> and reliable app for communication. Thanks."	trust
<b>awesome</b>	"It is <b>awesome</b> . I love it. Whatsapp is my favorite app."	satisfaction
	"It's always <b>awesome</b> . It deserves full rating ..."	hedonics
	"You have it because everyone has it. A 'smartphone' is defined by its capability to run this app! <b>Awesome</b> . Saved me a whole lot of money undoubtedly."	satisfaction
<b>communicate</b>	"Great app to <b>communicate</b> quickly and easily ..."	enjoyment,
	"A great way to <b>communicate</b> with friends and family. So clear without a hitch."	enjoyment
	"It's very practical. Great audio in the calls. Simply the most consistent form of <b>communicate</b> on the internet."	enjoyment
<b>super</b>	"This App has made my texting so much quicker and is <b>super</b> -fast sending pics and video's. My wife and I love it and text each other only on WhatsApp! Get it and you won't be sorry!"	hedonics
	" <b>Superb</b> application...user friendly ... just there should be some kind of indications of those who are online like we have in Facebook ... a green signal or something like that should be there so that we don't have to check that who are online. Other than that it's perfect."	hedonics

# Parameter Setting Problem in the Case of Practical Vehicle Routing Problems with Realistic Constraints

Emir Žunić

Info Studio d.o.o. Sarajevo and Faculty of  
Electrical Engineering, University of  
Sarajevo, Bosnia and Herzegovina  
Email: emir.zunic@infostudio.ba

Dženana Đonko

Faculty of Electrical Engineering,  
University of Sarajevo, Bosnia and  
Herzegovina  
Email: ddonko@etf.unsa.ba

**Abstract**—Vehicle Routing Problem (VRP) is the process of selection of the most favorable roads in a road network vehicle should move during the customer service, so as such, it is a generalization of problems of a commercial traveler. Most of the algorithms for successful solution of VRP problems are consisted of several controll parameters and constants, so this paper presents the data-driven prediction model for adjustment of the parameters based on historical data, especially for practical VRP problems with realistic constraints. The approach is consisted of four prediction models and decision making systems for comparing acquired results each of the used models.

## I. INTRODUCTION

THE problem of transport route optimization and optimal exploitation of the transport fleet has been explored and constantly improved for a long time. Vehicle routing problem is the name for the entire group of problems requiring the optimal route the transport vehicle or more of them (one vehicle can be used more than once during a single routing) can go around the specific number of customers (delivery points), starting from the central depot and returning there after the customer service. The optimal route is the one with the minimal cost of the road charges [1]. These optimization problems are becoming extremely complex considering a large number of customers. Additionally, those problems become a real challenge considering numerous constraints and facts such as time windows (TW) of customers, time of goods unloading, goods packaging into vehicles, predefined capacity and various vehicles, fixed and variable vehicle costs, etc. These indicators drastically increase the number of available approaches, models and algorithms which could be applied to a complex set of input data. The standard constraints VRP problems differ from are: the number of depots (one or more of them), maximum allowable timing or the length of the vehicle route, different vehicle capacities, customers' demands for delivery or collection of a certain amount of cargo during the service, time windows for beginning and finishing customer service as well as vehicle time windows.

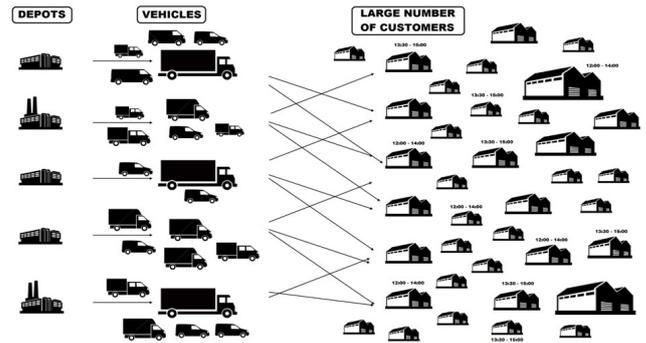


Fig. 1 Rich Vehicle Routing Problem

In a realistic surroundings, it is necessary to take into consideration a huge number of additional constraints, usually being the result of the specificity of the loading and/or unloading locations, specific business processes of the company performing the distribution (collection) or legal decisions and obligations (such as having a rest period for drivers). Such problems are often called Rich Vehicle Routing Problems (RVRP), and one illustrative example is shown on Fig. 1. There are two different approaches in solving RVRP. One of them is the exact algorithms with the aim of finding the optimal solution and proving it to be optimal. Another approach include approximate approaches (heuristic and metaheuristic), with the aim of finding the best possible solution, but without proving it to be optimal. Heuristic approaches and methods are commonly used in complex VRP problems. Neural Networks (NN) as well as Machine Learning (ML) methods have lately been used for the solution of complex VRP problems. Any of these approaches used for solving VRP problems also indicate there are constants and parameters of the algorithm, and their adjustment can make better or worse solution.

The next section presents the basic ways and approaches of adjusting the VRP algorithm parameters. The third section shows the access for adjustment of the control VRP algorithm parameters, consisting of several data transformations. Four algorithms are used for regression. Comparative analysis and discussion of the acquired results are presented in the fourth section. The final section depicts the conclusion of the paper, as well as the guidelines for future researches.

□ This work was supported by Info Studio d.o.o. Sarajevo

## II. APPROACHES TO ADJUSTMENT OF VRP ALGORITHM PARAMETERS

In most of the papers available in literature, there is the fact which points to every realistic VRP problem to be a bit different from its the most similar problem. Two groups of parameters (controlling ones) are the reason of that: (i) Some realistic constraints and constants of the input data, (ii) Constants of the used algorithm. Every company which requires their implementation of the VRP within their surroundings, has its own constraints, defined by the company business polices. Therefore, it is mentioned in literature that the constraints and restrictions in these kinds of problems are non-standard. In the paper [2], Lee describes in details one such problem with the possible solution of the particular example. With the development of modern technology, in the last few years, determination of the parameters based on the available information from GPS and/or GIS system, data on weather condition and forecasts, has also been performed. Several papers deal with the analysis of such systems, and especially interesting papers are [3] those with the data mining methods and techniques for determination of specific realistic constraints of the VRP problems, and [4] those using the predictions of time distances between knots for dynamic routing requirements for emergency vehicles.

One of the most interesting examples of the classical application of realistic and useful data is presented in paper [5], where the concept of data-driven solution of VRP problem is introduced for the first time. The additional phase is also mentioned for the first time in this paper, and it is used in examples that could be applied in realistic surroundings, and that is Human-Computer Interaction Mode (HCIM), which enables the end user to have the ability of manual modification of the suggested routes. No matter how the algorithm for solving VRP problems is considered perfect, there are always realistic situations that are impossible to predict and include in it, so the possibility of manual modification of the suggested routes in practical systems is of a great importance. Each of the analyzed approaches and algorithms for solving VRP problems is composed of specific constants and controlling parameters. Those parameters and constants are used for adjusting specific weight factors, punitive factors according to individual criteria depending on the importance of the very criterion on the final outcome of the realistic situation of vehicle routing, etc. In literature, this approach is defined as the Parameter Setting Problem (PSP). The most interesting paper on this subject was presented by Calvet et al. [6], describing the statistical approach for fine adjusting of the parameters for metaheuristic algorithms, which is also applied to VRP problem.

Analyzing the other available literature dealing with PSP problem, it can be concluded that these problems could be classified into two basic groups [7]: (i) *Parameter Control Strategies* (PCS), (ii) *Parameter Tuning Strategies* (PTS). In

papers [8]-[9], there is additional sub-group IPTS (Instance-Specific Parameter Tuning Strategies), which includes the characteristics of the instances being applied to. Although there are not many published scientific papers in available literature on the subject of PSP problems, it can be noticed that one of them most interesting applications of it is the improvement of certain segments of VRP, as well as the facilitating of solving VRP problems.

Battiti and Brunato [10] also presented an interesting paper on this subject in which they use the methods of machine learning in combination with statistical methods for fine adjustment of parameters of metaheuristic algorithms. They presented the model that could also be applied for parameters adjustment in other types of algorithms, and one of the most interesting examples is the application for the parameters of neural networks.

Some of the starting ideas for solving realistic VRP problems, and parameters (constants) settings problem are presented in paper [11], where the way of using GPS/GIS data for setting the attribute of the algorithms is presented. In paper [12], the cluster-based analysis and time-series prediction model for reducing the number of traffic accidents is presented.

## III. DATA-DRIVEN APPROACH FOR ADJUSTING THE CONTROL PARAMETERS OF THE VRP ALGORITHMS

There are three data sets VRP problem is consisted of: depot, vehicles and customers (users). At least one depot must be defined in the problem, and important information that should be collected for the depot is:

- *Location*: Address, Postal code and place, Geographic position (latitude and longitude)
- *Working hours*: Opening time, Closing time

If there are more depots distribution is made from, where customers can be served from any other depot, the problem can be modeled as a Multiple Depot Vehicle Routing Problem (MDVRP). But if customers are connected to a particular depot, it is necessary to model more individual VRP problems, for each depot and its buyers individually. In multiple depot problems, vehicles are usually the part of that depot. Basic information about the vehicle include:

- *Load space capacity*: Capacity [kg], Volume [m<sup>3</sup>], Number of pallet positions [pcs], Number of cargo units [pcs]
- Driver's working time
- Departure location
- Arrival location

The capacity of load space is limited by several criteria. For example, in distribution of goods, it is important to pay attention to limited capacity as well as the volume of load space to be sufficient for all the goods being transported. Certain types of goods can be light, but occupy a lot of

space, so it is necessary to fulfill both constraints. Goods are often put on pallets, so the capacity of vehicle can include the number of pallets possible to place into to load space. The simplest case is in distributing the goods of the same dimensions, so the capacity can be expressed by the maximum number of pieces that could be placed into the load space. Customers' data usually include:

- *Location*: Address, Postal code and place, Geographic position (latitude and longitude)
- *Order*: Weight [kg], Volume [m<sup>3</sup>], Usage of the pallets [%], Number of load units [pcs]
- *Time limits*: The earliest discharge time, The latest discharge time, Estimated discharge duration

Every implemented algorithm solving the realistic VRP problem is consisted of control parameters and constants. The value of these parameters and constants could be adjusted based on historical data, and the model depicting the way of that adjustment is presented in Fig. 2.

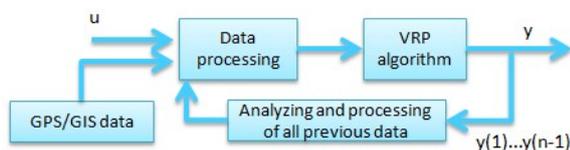


Fig. 2 Approach for adjusting the control parameters based on historical data

The primary goal of realistic VRP problems is to fulfill all constraints. Some of these control parameters are adjusted as described below, using some of the prediction methods and algorithms which will make conclusions based on historical data. On the basis of all the parameters that could have an impact to the final outcome of the solution of VRP problems, several basic ones are sorted out and stored for every route during the testing and production use, since the historical data of practical implementation of VRP algorithms have been used in several largest distribution companies in Bosnia and Herzegovina, dealing with product distribution from their depots to delivery points (shops, supermarkets, etc.). It resulted in creation of a knowledge base being enriched every day, and the main goal is the adjustment of control parameters and constants on the basis of historical data, which can later become a part of any implemented VRP algorithm. Attributes being sorted out as those affecting the routes are:

- The number of customers
- The number of available vehicles
- The number of available different types of vehicles
- The number of towns
- The total number of restrictions where a customer can't be served by a certain vehicle
- Total number of ordered articles [pcs]
- Total volume of ordered articles [m<sup>3</sup>]

- Total weight of all the articles [kg]
- Total duration of TW of all the customers [min]
- Are all the restrictions fulfilled (1: yes, 0: no)

The target attributes affecting the given routes and total cost presenting the control parameters of VRP algorithms in the case of distribution companies are:

- *ToleranceWeight*
- *ToleranceVolume*
- *PenaltyDelay*
- *PenaltyCustomersVehicles*
- *CostIncreasing*
- *PenaltyVolumePercentage*
- *PenaltyWeightPercentage*

During the implementation of VRP algorithms, the vehicle can be allowed to be overloaded in weight or volume, for the values of *ToleranceWeight* and *ToleranceVolume*. Algorithms for solving VRP problems allow the vehicle to be delayed for customer (to arrive outside of its time window). The violation of this parameter is presented by the *PenaltyDelay*. If the VRP algorithm is adjusted for solving Site-Dependent Vehicle Routing Problem (SDVRP), the attribute *PenaltyCustomersVehicles* is used for penalization of rules violations, where the customer can't be served by a particular vehicle. The constant used in increasing the cost, depending on the weight vehicle transports, is presented with the *CostIncreasing*. The constants penalize reloading of vehicles by weight or volume are *PenaltyWeightPercentage* and *PenaltyVolumePercentage*, and they present the cost increasing when the weight/volume of the vehicle is reloaded by 100%. Each of these parameters is determined independently. Firstly, there is a data preprocessing, by excluding only those historical data where all the constraints are fulfilled (value 1 in the column). Then, the removal of redundant attributes was performed by using the Transform Option. Using the Attribute Importance option, the determination of input attribute importance was made for every target attribute. Minimum Descriptor Length (MDL) algorithm was used for attribute importance determination. Before that, the normalization of attribute had been performed in a way the volume of the attribute was put to one decimal place.

After all the preprocessings and preparations of input values, the proposed model was created for determining the target attributes. Model for one attribute is shown in Fig. 3.



Fig. 3 Prediction model for one attribute

Four regression algorithms were used, and their results are compared: (1) Generalized Linear Models (GLM); (2) Support Vector Machine (SVM); (3) Decision Tree (DT); (4) Naive Bayes (NB).

The advantage of SVM over other methods is providing better predictions in unseen test data, providing unique optimal solutions for the problem and the existence of less parameter for optimization compared to other methods. The speed of performance is not crucial for the problem wanted to be applied on, so the lack of SVM regression method can be ignored. GLM regression algorithm is chosen because it represents the generalization of the linear regression and is often used in cases where output variables do not have normal distribution. Since the input data point to linear dependence, the GLM choice of the regression algorithm was a logical choice. Basic advantages of a Decision Tree method are: the possibility of generating comprehensive models, relatively small requirements for computer resources (time and memory) and precise importance of some attributes for the specific problem, as well as vast availability of software solution. The lack of Decision Trees is their instability, because small fluctuations in data sample can result in huge variations in assigned classifications. The advantage of the Naive Bayes classifier is its robustness to errors obtained in data collecting or missing attribute value in training session. Errors do not have a lot of impact on probabilities since they are of average value, while missing data are simply ignored during the calculating of probability. Also, the Naive Bayes classifier is robust to irrelevant attributes, too.

#### IV. RESULTS DISCUSSION

As previously mentioned, for each of the control parameters, the independent model was created with four regression algorithms: GLM, SVM, Decision Tree and Naive Bayes algorithm. After the results were obtained, for each of the parameters, the Decision Support System (DSS) was created which, on the basis of regression results, chose the predicted value of the algorithm that had higher Predictive Confidence. In order to enrich the control parameters knowledge base, VRP productive algorithms were run for more than 10.000 times, with all the constraints fulfilled for many different days and input parameters. Data with testing and validating of prediction models are provided at the 4TU Research Data Center [13], to be available to other researchers for their works and eventual comparison of results. For each of the control parameters, there was a result comparison of confidence/accuracy - Predictive Confidence [%], which was presented for each of the input data set in Table I.

As seen from the presented results, it is not difficult to conclude that the SVM always provided better predictive results for each of the control parameters compared to GLM algorithm (Table I – left), while the prediction models based on the Decision Tree and Naive Bayes algorithms, always showed much worse results (Table I – right), and therefore, the Decision Support System (DSS) preferred the prediction results of the SVM algorithm. Analyzing the Table I in details, it is easy to conclude that the lack of Decision Tree is

their instability, because small fluctuations may result in large variations in the assigned classifications, which was the case. The lack of the Naive Bayes prediction algorithm is presumably the independence of attributes, which make this classifier delicate to correlated attributes. Attributes in strong correlation can degrade performances of the classifiers, which can be solved by removing certain attributes, which is also the case in this example.

TABLE I - COMPARATIVE RESULTS OF USED REGRESSION GLM, SVM, DT AND NB MODELS – PREDICTIVE CONFIDENCE [%]

Parameter	GLM	SVM	DT	NB
<i>ToleranceWeight</i>	91.766	96.123	81.336	84.323
<i>ToleranceVolume</i>	81.995	90.201	79.928	80.905
<i>PenaltyDelay</i>	84.956	89.551	81.555	82.007
<i>PenaltyCustomersVehicles</i>	92.031	96.439	82.309	85.314
<i>CostIncreasing</i>	81.276	89.996	78.998	80.229
<i>PenaltyVolumePercentage</i>	89.133	91.853	83.892	84.934
<i>PenaltyWeightPercentage</i>	90.006	92.698	84.801	84.956

Implemented Attribute Importance segment at the target values in the prediction model enables to determine the importance of each of the input attributes on the target control parameter. The average importance of the input parameters on the input control variables, as well as their order, is shown in Table II. Analyzing the average value of input attribute influences for each of the control parameters, it is concluded that the input parameters affect the output prediction control parameters in that order shown in Table II. Such results were expected because routings made on realistic data were extremely complex with strict constraints, while the number of available vehicles was very small. The fact is that out of 8 available vehicles for routing, seven of them were of different type, which significantly affects the result and complexity of algorithm implementation. It is easy to conclude that these parameters are the most important for adjusting the algorithm control parameters. Also, the customers' time windows are significant for the control parameters, which affect the complexity of finding a solution.

TABLE II - ATTRIBUTE IMPORTANCE RESULTS

Input Parameter	Importance Number	Importance Weight
<i>The number of available different types of vehicles</i>	1	0.871
<i>The number of available vehicles</i>	2	0.861
<i>Total duration of time windows of all the customers [min]</i>	3	0.796
<i>Total number of ordered articles [pcs]</i>	4	0.758
<i>The number of customers</i>	5	0.747
<i>Total volume of ordered articles [m<sup>3</sup>]</i>	6	0.741
<i>The number of towns</i>	7	0.727
<i>Total weight of all the articles [kg]</i>	8	0.582
<i>The total number of restrictions where a customer can't be served by a certain vehicle</i>	9	0.303

According to Table II, the parameter with the least significance for adjusting the value of control parameters is the number of constraints the customer can't be served by a vehicle. The very number is presented in the form of the summary indicator, but if it were presented in terms of ratio

of the customer and the number of vehicles that could serve him, the significance of that parameter would certainly be much greater, even the most significant. For each of the control parameters, the results are graphically displayed, which is example of the *ToleranceWeight* parameter shown in Fig. 4 (left). The important indicator in these analyses is the Mean Absolute Error (MAE) prediction. Results of the MAE for the example of *ToleranceWeight* parameter are shown in Fig. 4 (right).



Fig. 4 Predictive confidence [%] and Mean Absolute Error results: SVM and GLM

It is also possible to observe the comparison of Residual (Residual is the difference between expected and predicted value of the dependent variable) for each of the control parameters. The example of the comparison of the *ToleranceWeight* parameter is shown in Fig. 5.

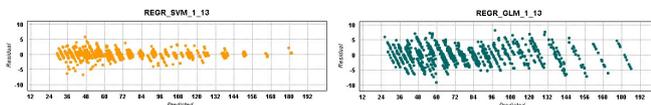


Fig. 5 Residual comparisons: SVM (left) and GLM (right)

Figure 6 shows that for the each attribute except the Predictive Confidence indicator, it is possible to obtain many other parameters (primarily refers to the prediction value errors), which enables to make comparisons and select the model which satisfies the needs and expectations more.

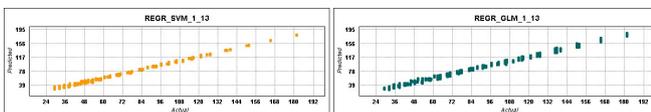


Fig. 6 Comparison of realistic and predicted values – SVM and GLM

It is also simple to make comparison of the actual and predicted value, and make the same conclusion that the SVM algorithm presented much better results compared to the remaining 3 prediction algorithms already used. Thus, every analysis points to the fact that the SVM algorithm is much more superior than other algorithms already used, as well as methods for the purpose of adjusting the control parameters of the practically applicable algorithms for solving the complex VRP problems.

## V. CONCLUSION

Regardless of what approach (exact or heuristics) is used for successful solving the VRP problem, most of the proposed algorithms are consisted of constants in control parameters whose adjustment give better or worse solutions. This paper presents the innovative approach of adjustment of these parameters on the basis of available historical data by

using the prediction models. Four of the prediction models were used, where SVM algorithm proved to have much better and more superior results for all the tests compared to other prediction models. For each of the analyzed control parameters in the case of SVM model, the predictive accuracy was over 90%. The advantage of SVM over other methods being used is providing better predictions in unseen test data, providing unique optimal solutions for the training problem, and the existence of less optimization parameters compared to other methods. The execution speed is not crucial for the problem, so the lack of SVM regression method can be neglected in this case.

Guidelines for the future researches in this area would include the application of neural networks for determination of the values of the VRP algorithm control parameters, or even the prediction based on time series. Surely, the progress should be realized by even more input set of realistic data in this segment of the proposed approach.

## REFERENCES

- [1] Dantzig, G. B., Ramser, J. H. 1959. The truck dispatching problem. *Management science*. 6(1):80-91, <https://doi.org/10.1287/mnsc.6.1.80>
- [2] Lee, W. L. 2013. Real-Life Vehicle Routing with Non-Standard Constraints. *Proceedings of the World Congress on Engineering (WCE)*. I:432-437
- [3] Hu, X., Huang, M., Zeng, A. 2007. An intelligent solution system for a vehicle routing problem in urban distribution. *International Journal of Innovative Computing, Information and Control*. 3:189-198
- [4] Musolino, G., Rindone, C., Polimeni, A., Vitetta, A. 2013. Travel Time Forecasting and Dynamic Routes Design for Emergency Vehicles. *Procedia - Social and Behavioral Sciences*. 87:193-202, <https://doi.org/10.1016/j.sbspro.2013.10.603>
- [5] Fu, C., Wang, H. 2010. The solving strategy for the real-world vehicle routing problem. *3rd International Congress on Image and Signal Processing*. 3182-3185, <https://doi.org/10.1109/CISP.2010.5647968>
- [6] Calvet, L., Juan, A. A., Serrat, C., Ries, J. 2016. A statistical learning based approach for parameter fine-tuning of metaheuristics. *SORT - Statistics and Operations Research Transactions*. 40(1):201-240
- [7] Birattari, M., Kacprzyk, J. 2009. *Tuning metaheuristics: A machine learning perspective*. Springer, Vol. 197. ISBN: 3642004822 9783642004827
- [8] Montero, E., Riff, M. C., Neveu, B. 2014. A beginner's guide to tuning methods. *Applied Soft Computing*. 17:39-51, <https://doi.org/10.1016/j.asoc.2013.12.017>
- [9] Ries, J., Beullens, P., Salt, D. 2012. Instance-specific multi-objective parameter tuning based on fuzzy logic. *EJOR*. Elsevier. 218:305-315, <https://doi.org/10.1016/j.ejor.2011.10.024>
- [10] Battiti, R., Brunato, M. 2010. *Reactive Search Optimization: Learning While Optimizing*. Handbook of Metaheuristics. International Series in Operations Research & Management Science. 543-571, [https://doi.org/10.1007/978-1-4419-1665-5\\_18](https://doi.org/10.1007/978-1-4419-1665-5_18)
- [11] Žunić, E., Hindija, H., Beširević, A., Hodžić, K., Delalić, S. 2018. Improving Performance of Vehicle Routing Algorithms using GPS Data. *14th Symposium on Neural Networks and Applications (NEUREL)*. 1-4. <https://doi.org/10.1109/neurel.2018.8586982>
- [12] Žunić, E., Djedović, A., Đonko, D. 2017. Cluster-based analysis and time-series prediction model for reducing the number of traffic accidents. *International Symposium ELMAR*. 25-29, <https://doi.org/10.23919/ELMAR.2017.8124427>
- [13] Žunić, E. (Emir). 2018. Real-world VRP data with realistic non-standard constraints - parameter setting problem regression input data. *4TU.Centre for Research Data*. Dataset. Available at: <https://doi.org/10.4121/uuid:97006624-d6a3-4a29-bffa-e8daf60699d8>



# Joint 39<sup>th</sup> IEEE Software Engineering Workshop and 6th International Workshop on Cyber-Physical Systems

**T**HE IEEE Software Engineering Workshop (SEW) is the oldest Software Engineering event in the world, dating back to 1969. The workshop was originally run as the NASA Software Engineering Workshop and focused on software engineering issues relevant to NASA and the space industry. After the 25<sup>th</sup> edition, it became the NASA/IEEE Software Engineering Workshop and expanded its remit to address many more areas of software engineering with emphasis on practical issues, industrial experience and case studies in addition to traditional technical papers. Since its 31<sup>st</sup> edition, it has been sponsored by IEEE and has continued to broaden its areas of interest.

One such extremely hot new area are Cyber-physical Systems (CPS), which encompass the investigation of approaches related to the development and use of modern software systems interfacing with real world and controlling their surroundings. CPS are physical and engineering systems closely integrated with their typically networked environment. Modern airplanes, automobiles, or medical devices are practically networks of computers. Sensors, robots, and intelligent devices are abundant. Human life depends on them. CPS systems transform how people interact with the physical world just like the Internet transformed how people interact with one another.

The joint workshop aims to bring together all those researchers with an interest in software engineering, both with CPS and broader focus. Traditionally, these workshops attract industrial and government practitioners and academics pursuing the advancement of software engineering principles, techniques and practices. This joint edition will also provide a forum for reporting on past experiences, for describing new and emerging results and approaches, and for exchanging ideas on best practice and future directions.

## TOPICS

The workshop aims to bring together all those with an interest in software engineering. Traditionally, the workshop attracts industrial and government practitioners and academics pursuing the advancement of software engineering principles, techniques and practice. The workshop provides a forum for reporting on past experiences, for describing new and emerging results and approaches, and for exchanging ideas on best practice and future directions.

Topics of interest include, but are not limited to:

- Experiments and experience reports

- Software quality assurance and metrics
- Formal methods and formal approaches to software development
- Software engineering processes and process improvement
- Agile and lean methods
- Requirements engineering
- Software architectures
- Design methodologies
- Validation and verification
- Software maintenance, reuse, and legacy systems
- Agent-based software systems
- Self-managing systems
- New approaches to software engineering (e.g., search based software engineering)
- Software engineering issues in cyber-physical systems
- Real-time software engineering
- Safety assurance & certification
- Software security
- Embedded control systems and networks
- Software aspects of the Internet of Things
- Software engineering education, laboratories and pedagogy
- Software engineering for social media

## EVENT CHAIRS

- **Bowen, Jonathan**, Museophile Ltd., United Kingdom
- **Hinchey, Mike**(Lead Chair), Lero-the Irish Software Engineering Research Centre, Ireland
- **Szmuc, Tomasz**, AGH University of Science and Technology, Poland
- **Zalewski, Janusz**, Florida Gulf Coast University, United States

## PROGRAM COMMITTEE

- **Ait Aneur, Yamine**, IRIT/INPT-ENSEEIH, France
- **Banach, Richard**, University of Manchester, United Kingdom
- **Cicirelli, Franco**, Universita della Calabria, Italy
- **Ehrenberger, Wolfgang**, Hochschule Fulda, Germany
- **Forbrig, Peter**, University of Rostock, Germany, Germany
- **Friesel, Anna**, Technical University of Denmark, Denmark
- **Gomes, Luis**, Universidade Nova de Lisboa, Portugal

- **Gracanin, Denis**, Virginia Tech, United States
- **Havelund, Klaus**, Jet Propulsion Laboratory, California Institute of Technology, United States
- **Hsiao, Michael**, Virginia Tech, United States
- **Letia, Tiberiu**, Technical University of Cluj-Napoca, Romania
- **Li, Jianwen**, Iowa State University, United States
- **Minchev, Zlatogor**, Bulgarian Academy of Sciences, Bulgaria
- **Nesi, Paolo**, DSI-DISIT, University of Florence, Italy
- **Obermaisser, Roman**, Universität Siegen, Germany
- **Pu, Geguang**, East China Normal University
- **Pullum, Laura**, Oak Ridge National Laboratory, United States
- **Qin, Shengchao**, Teesside University, United Kingdom
- **Reeves, Steve**, University of Waikato, New Zealand
- **Roman, Dumitru**, SINTEF / University of Oslo, Norway
- **Rysavy, Ondrej**, Brno University of Technology, Czech Republic
- **Sanden, Bo**, Colorado Technical University, United States
- **Sekerinski, Emil**, McMaster University, Canada
- **Selic, Bran**, Simula Research Lab, Norway
- **Sojka, Michal**, Czech Technical University, Czech Republic
- **Sun, Jing**, The University of Auckland, New Zealand
- **Trybus, Leszek**, Rzeszow University of Technology, Poland
- **van Katwijk, Jan**, Delft University of Technology, The Netherlands
- **Vardanega, Tullio**, University of Padova, Italy
- **Velev, Miroslav**, Aries Design Automation, United States
- **Vilkomir, Sergiy**, East Carolina University, United States
- **Waeselynck, Hélène**, LAAS-CNRS Toulouse, France
- **Zhu, Huibiao**, Software Engineering Institute—East China Normal University

# Handling of Categorical Data in Software Development Effort Estimation: A Systematic Mapping Study

Fatima Azzahra Amazal

LabSIV, Department of Computer Science,  
Faculty of Science, Ibn Zohr University, BP  
8106, 80000 Agadir, Morocco  
Email: amazal.fatimaazzahra@gmail.com

Ali Idri

Software Projects Management Research Team,  
ENSIAS, Mohamed V University, Madinat Al  
Irfane, 10100 Rabat, Morocco  
Email: idri@ensias.ma

**Abstract**—Producing reliable and accurate estimates of software effort remains a difficult task in software project management, especially at the early stages of the software life cycle where the information available is more categorical than numerical. In this paper, we conducted a systematic mapping study of papers dealing with categorical data in software development effort estimation. In total, 27 papers were identified from 1997 to January 2019. The selected studies were analyzed and classified according to eight criteria: publication channels, year of publication, research approach, contribution type, SDEE technique, Technique used to handle categorical data, types of categorical data and datasets used. The results showed that most of the selected papers investigate the use of both nominal and ordinal data. Furthermore, Euclidean distance, fuzzy logic, and fuzzy clustering techniques were the most used techniques to handle categorical data using analogy. Using regression, most papers employed ANOVA and combination of categories.

## I. INTRODUCTION

THE competitiveness of software companies relies on the successful management of their software projects. One of the most important and difficult tasks in software project management is how to accurately estimate the effort needed to develop a software product. This task is known as software development effort estimation (SDEE). Delivering reliable and accurate estimates remains a challenging objective for software companies due to several factors including the human factor, the variety of software projects, the inherent uncertainty of feature measurement, and the diversity of development environments [1]. In attempt to get accurate predictions, various SDEE techniques have been proposed. These techniques fall into three main types [2]: parametric models [3], [4], machine learning (ML) models [5]-[10] and expert judgment [11].

SDEE techniques build their predictions based on a set of attributes (also called features or cost drivers) that characterize software projects [12], [13]. Most of these techniques derive their predictions based on numerical attributes. However, the information available at the early stages of the software life cycle is more categorical than numerical. Furthermore, the datasets used to build and validate SDEE models

involve a high number of categorical data. For example, in COCOMO'81 dataset [14], 15 attributes out of 17 are measured on a scale composed of six categories: very low, low, nominal, high, very high, and extra high. Another example is the International Software Benchmarking Standards Group (ISBSG) dataset [15], in which numerous attributes such as programming language, application type and development platform are measured on a nominal scale.

Categorical attributes may be measured on a nominal or ordinal scale. The nominal scale type allows the classification of entities into different categories [16], for example, primary programming language may be classified into five categories: Visual basic, C, Cobol, Visual C++, Oracle. Unlike the nominal scale type in which there is no order between the categories of entities, the ordinal scale type enables ranking the categories in a specific order [16]. An example of ordinal attributes is the application experience which may be measured as: 'low', 'nominal', 'high', and 'very high'. To deal with this kind of attributes, different approaches were used in SDEE literature [17]-[21].

In this paper, a Systematic Mapping Study (SMS) is performed to investigate the use of categorical data to estimate software development effort. As pointed out in [22], a systematic map is a method that concentrates on building a classification scheme and categorizing primary research studies in a specific domain with respect to a set of defined categories. Thus, it provides a common starting point for many researchers [23]. To the best of the authors' knowledge, no systematic mapping study has been carried out with focus on how to handle categorical data in SDEE.

This SMS aims to: 1) identify the existing SDEE papers dealing with categorical data and published from 1997 to January 2019; and 2) analyze and classify the selected papers according to 8 criteria: publication channels, year of publication, research approach, contribution type, SDEE technique, Technique used to handle categorical data, types of categorical data and datasets used.

This paper is structured as follows: Section II presents the research methodology adopted to carry out this SMS.

Section III, reports the results of the mapping study. Section IV presents the implications for research and practice. Conclusions and future work are presented in Section V.

## II. RESEARCH METHODOLOGY

In this study, the systematic mapping process suggested by Kitchenham and Charters [24] is used. According to Kitchenham, a mapping study aims to identify the research trends related to a specific topic and classify research works with respect to a set of defined criteria [22], [24]. The mapping process used comprises the following five steps: (1) define the mapping questions, (2) conduct an exhaustive search for candidate papers, (3) select studies, (4) extract data, and (5) summarize data. Each of these steps is described next.

### A. Mapping questions

Eight mapping questions (MQs) were formulated in this mapping study. Table I shows the MQs as well as their main motivations.

### B. Search Strategy

The aim of this step is to find the relevant SDEE papers that address the MQs listed in table I. To perform the search, four electronic databases were used: ACM Digital library, IEEE Xplore, Science Direct and Google Scholar. These libraries were chosen since they were used in previous systematic maps and reviews in SDEE to conduct the search for candidate papers [5], [25], [26]. All searches were restricted to the studies published between 1997 and January 2019.

TABLE I. MAPPING QUESTIONS

ID	Mapping Question	Motivation
MQ1	Which publication sources are the main targets for SDEE papers dealing with categorical data?	To identify the main sources where SDEE studies with focus on categorical data can be found.
MQ2	How has the frequency of handling categorical data in SDEE papers changed over time?	To investigate the publication trends of SDEE studies dealing with categorical data over time.
MQ3	What are the research approaches of the selected papers?	To discover the research approaches used by SDEE studies with focus on categorical data.
MQ4	What are the contribution types of the selected papers?	To explore the contribution types of SDEE papers dealing with categorical data.
MQ5	Which technique investigates the most the use of categorical data in SDEE?	To identify the SDEE techniques that handle the most categorical data.
MQ6	How categorical data are handled in SDEE?	To determine the different ways of handling categorical data in SDEE.
MQ7	What are the most investigated types of categorical data in SDEE?	To identify the types of categorical data that are the most investigated in SDEE.
MQ8	What are the datasets used for validation?	To explore the datasets used in the selected papers as well as the Percentage of categorical features

	used in the experiments.
--	--------------------------

To carry out the search using the four databases, a search string was defined. To do so, we derived the main terms based on the MQs. Then, we identified all alternative spellings and synonyms of the major terms. The Boolean operators OR and AND were used to combine the main terms [25], [26]. The final search string was formulated as follows:

(software OR system OR application OR product OR project OR development) AND (effort OR cost) AND (estimat\* OR predict\* OR assess\*) AND (categorical OR nominal OR ordinal OR "non-quantitative") AND (feature OR attribute OR data OR "cost driver").

To ensure that no relevant paper was missed, we adopted a search process of two stages. In the first stage, we performed the search in the four electronic databases using the above search string to identify the set of candidate papers. In the second stage, we applied the inclusion and exclusion criteria on each of the candidate papers based on title, abstract, and keywords to decide on its relevance to our study. If necessary, the full paper was examined. The reference list of each of the relevant papers was scanned to check whether a SDEE study with focus on categorical data was leaved out in the first stage.

### C. Study Selection

The purpose of this step was to select the papers that are relevant to our SMS (i.e., papers that addressed the MQs). To achieve this, a set of inclusion and exclusion criteria were applied on each of the candidate papers by each of the authors of this study to decide whether it should be retained or discarded.

#### Inclusion criteria

- ✓ Studies with focus on how to handle categorical data to estimate software effort
- ✓ Studies in which a technique is proposed or extended and which enables software effort estimation using categorical data or a mixture of numerical and categorical data
- ✓ Studies comparing different techniques that handle categorical data

#### Exclusion criteria:

- ✓ SDEE studies in which categorical features are not handled or discarded
- ✓ SDEE studies for which the main objective is not deal with categorical data and which use only transformation to dummy variables
- ✓ SDEE studies that fuzzify numerical inputs to get linguistic values without dealing with categorical inputs
- ✓ SDEE studies with focus on missing categorical data
- ✓ Duplicate publications of the same paper (In this case, only the most complete study is included)
- ✓ Studies estimating maintenance or testing effort

Using the above criteria, the two researchers independently evaluate the candidate papers. Based on the title and abstract (if necessary full text), a researcher might categorize a candidate paper as "include", "Exclude", or "Uncertain". A paper that was categorized as "Include" ("Exclude") by both researchers was retained (discarded); otherwise, the paper was discussed until an agreement was reached.

**D. Data Extraction Strategy and Synthesis Method**

Each of the selected papers was examined by both authors to extract the data necessary to answer the mapping questions of table I. To this end, a data extraction form was used and completed by both authors for each selected paper. Table II shows the data extraction form used in our mapping study.

The extracted data were, then, synthesized and summarized with respect to each MQ. To achieve this, a narrative synthesis approach was used. We also used some visualization charts such as pie charts and bubble plots to improve the presentation of the results obtained and facilitate their interpretation.

TABLE II. DATA EXTRACTION FORM

<b>Data extractor</b>
<b>Paper identifier</b>
<b>Author(s) name(s)</b>
<b>Article title</b>
(MQ1) Publication Channel
(MQ2) Publication year
(MQ3) Research approach (History-based evaluation, solution proposal, case study, theory, review, survey, other)
(MQ4) Contribution type (Technique, tool, comparison, validation, metric, model, framework)
(MQ5) SDEE Techniques used in the paper
(MQ6) Technique used to handle categorical data
(MQ7) Types of categorical data used in the study
(MQ8) Datasets used

**III. RESULTS AND DISCUSSION**

This section presents and discusses the results of our systematic mapping related to the questions of table I.

**A. Overview of the selected studies**

The results of the selection process are shown in Fig. 1. As can be seen, 1226 candidate papers were retrieved by applying the search string described previously on the four electronic databases. Afterward, the inclusion and exclusion criteria were used to evaluate each of the candidate papers and decide whether it should be retained or discarded. The evaluation was based on the title, abstract, keywords, and full text of the candidate papers. This process resulted in 27 relevant papers. No additional relevant studies were identified by checking the reference lists of the selected studies.

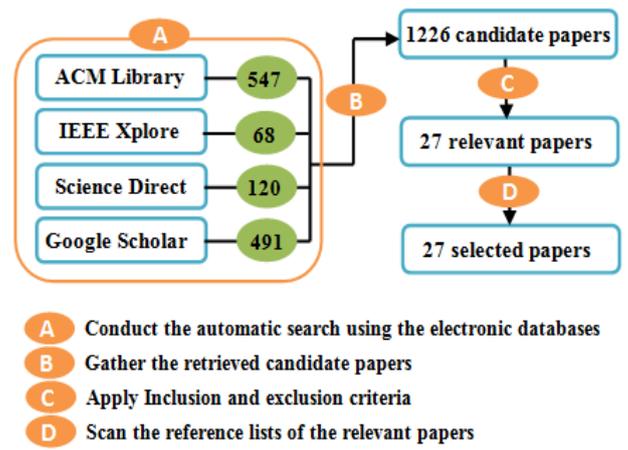


Fig. 1 Results of selection process

**B. Publications Channels (MQ1)**

We identified two main publication channels in which the selected studies were published: journals and conferences. Specifically, among the 27 selected papers, 15 (55.56%) papers appeared in journals and 12 (44.44%) papers were presented at conferences. Tables III and IV shows the publication sources of the papers identified in journals and conferences respectively. The number of studies per publication source is given in the second column of each table. Three journals were identified with 2 or more papers dealing with categorical data in SDEE: Empirical Software Engineering, Information Software Technology, and IEEE Transactions on Software Engineering. Only one conference was identified with 2 papers: International Conference on Predictive Models in Software Engineering (PROMISE). The remaining sources (journals and conferences) were used once to publish SDEE studies with focus on categorical data.

TABLE III. PUBLICATION SOURCES OF JOURNAL PAPERS

Publication venue	# of studies
Empirical Software Engineering	4
Information and Software Technology	3
IEEE Transactions on Software Engineering	2
The Journal of Systems and Software	1
International Journal of Intelligent Systems	1
International Journal of Computer Science and Engineering Survey	1
Software Quality Journal	1
Journal of Information Science and Engineering	1
IEEE Access	1

TABLE IV. PUBLICATION SOURCES OF CONFERENCE PAPERS

Publication venue	# of studies
International Conference on Predictive Models in Software Engineering	2
International Conference on Software Engineering Research, Management and Applications	1
Asia-Pacific Software Engineering Conference	1

Software Metrics Symposium	1
International Conference on Computer and Information Technology	1
International Conference on Software Engineering	1
International Conference on Computer Science and Automation Engineering	1
International Symposium on Software Metrics	1
International Conference on Enterprise Information Systems	1
International Conference on Communications, Circuits and Systems and West Sino Expositions	1
Empirical Software Engineering and Measurement	1

### C. Publications Trends (MQ2)

To get a global picture of the publication trends of SDEE papers dealing with categorical data, we analyzed the distribution of the selected studies over time. Fig. 2 shows the number of papers per year from 1997 to January 2019. As can be seen, the publication of SDEE papers with focus on categorical data is characterized by discontinuity. In fact, no paper was identified in some specific years (1998, 2000, 2003, 2005, 2014, 2017, 2018). Handling categorical data in SDEE has gained research interest in the period 2008-2013 (59% of the selected papers). Outside this period, poor number of studies was identified (not more than one paper per year except for 2001).

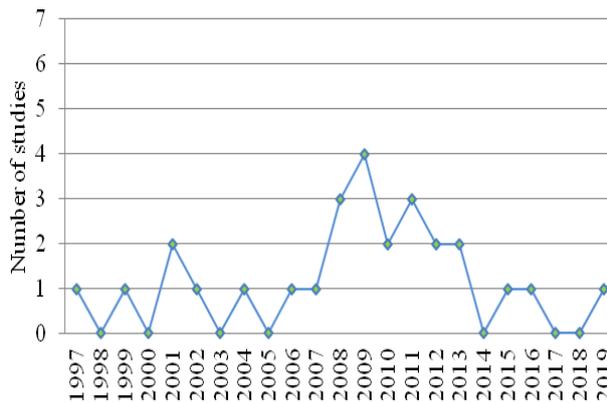


Fig. 2 Publication trends of the selected studies

### D. Research approaches (MQ3) and contribution types (MQ4)

As shown in Fig. 3, two main research approaches were used in the selected papers: solution proposal, and history-based evaluation. The solution proposal approach was adopted by 85% of the selected studies. Among them, 91% (21 out of 23) proposed new techniques, 4% (1 out of 23) proposed a new framework and 4% investigated the use of a new metric. Note that, all selected studies were included in the history-based evaluation approach. Among them, 15% (4 out of 27) performed a comparison of various SDEE techniques using datasets with mixed numerical and categorical data. The remaining papers used historical

datasets to assess the performance of their proposed approaches.

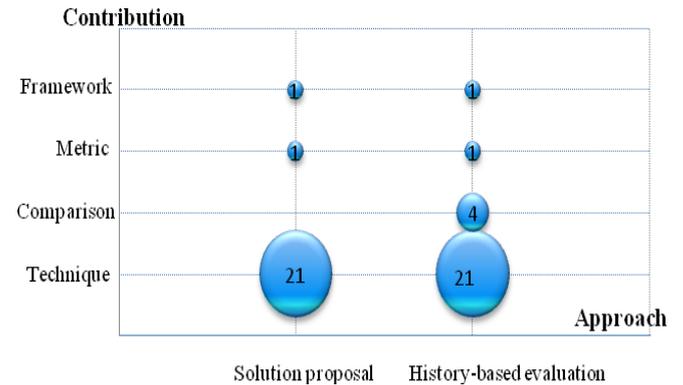


Fig. 3 Research approaches used in the selected studies and their contribution type

### E. SDEE Techniques investigating categorical data (MQ5)

Various approaches were used in the selected papers to estimate software effort using a mixture of numerical and categorical data. Table V shows the techniques used as well as the number of studies in which they were applied. Case based-reasoning (CBR), Regression (SR), Fuzzy Logic (FL), and Classification and Regression Trees (CART) were the techniques that investigate the most the use of categorical data in software effort estimation. Most of these techniques were not used alone. They were combined with each other to improve their prediction accuracy and to get accurate estimates. Specifically, 59% (16 out of 27) of the selected papers used a combination of two or more techniques to predict software effort whereas 41% employed a single technique.

TABLE V. TECHNIQUES USED IN THE SELECTED PAPERS

Technique used	# of studies	Studies
Case based-reasoning (CBR)	15	S3, S6, S7, S8, S9, S10, S12, S13, S15, S16, S17, S18, S19, S20, S24
Regression (SR)	9	S1, S4, S12, S16, S20, S22, S25, S26, S27
Fuzzy Logic (FL)	7	S2, S3, S8, S9, S14, S15, S21
Classification and Regression Trees (CART)	5	S11, S12, S14, S16, S21
Model Tree (MT)	2	S5, S7
Artificial Neural Networks (ANN)	2	S19, S26
Grey Relational Analysis (GRA)	2	S8, S9
Stepwise ANOVA	2	S12, S16
Bees Algorithm (BA)	1	S5
Kendall's Row-wise Rank Correlation	1	S6

(CRRC)		
Particle Swarm Optimization (PSO)	1	S10
Association Rules (AR)	1	S11
Mantel's Correlation (MC)	1	S17
Collaborative Filtering (CF)	1	S18
Genetic Programming (GP)	1	S23
IFPUG	1	S25

TABLE VI. CATEGORICAL (NOMINAL AND ORDINAL) DATA HANDLING

Categorical data handling	SDEE Technique	Studies
Euclidean distance	CBR	S6, S7, S10, S12, S16, S17, S19, S24
Combining categories / ANOVA	Regression	S4, S12, S16, S20, S22, S26
Classification by DT	Decision trees	S5, S11, S12, S14, S16, S21
Fuzzy logic	CBR / DT	S2, S3, S13, S15, S14
Fuzzy Clustering technique	CBR	S3, S13, S15
Quantification of data	Regression	S4
Grey Relational Coefficient	CBR	S8
Manhattan distance	CBR	S10
Local similarity	CBR	S18
Grammar Guided Genetic Programming	Genetic Programming	S23 [44]

*F. Handling of categorical data in SDEE (MQ6)*

To deal with categorical data, different techniques were applied depending on their type (nominal or ordinal) as well as the SDEE technique in which they were used. Table VI shows how both nominal and ordinal data were handled in the selected SDEE studies. Note that, some studies used the term 'Categorical' without specifying the exact data type (nominal or ordinal). As shown in table VI, using CBR, Euclidean distance is the most used metric to assess the similarity between two projects that are described by a mixture of numerical and categorical data [9], [27]-[33]. Fuzzy logic, and fuzzy clustering techniques were also used in many CBR/DT works to deal with categorical data [10], [17], [20], [34], [35]. Using regression, most papers employed one-way Analysis of Variance (ANOVA) and recorded categorical variables into new ones with fewer categories [2], [18], [30], [31], [36], [37]. Other studies employed classification and regression trees to handle categorical data [30], [31], [35], [38]-[40].

The above-mentioned techniques were applied to handle both nominal and ordinal data. Other techniques to deal with categorical data were identified depending on whether they are measured on a nominal or ordinal scale. Table VII shows how nominal data were handled in the selected papers. Using regression, four techniques were identified: Transformation to dummy variables, dataset segmentation, interaction, and use of a hierarchical linear model [30], [31], [36], [41], [42]. Using CBR, the equality distance was used to assess the similarity between projects that are described by nominal features [1], [36]. Regarding ordinal data, they were handled as if they were measured using an interval scale or converted to numerical values using regression [30], [43]. Using CBR, they were treated as interval scaled or handled using Grow's formula [1], [36] (see table VIII).

It is worth noting that, when investigating the use of categorical data in the selected papers, we found that some CBR works used categorical data not only to measure the similarity between software projects using Euclidean distance but also: 1) to adjust estimation by analogy; 2) to identify whether a categorical attribute is appropriate to yield predictions or 3) for feature weighting (see table IX).

Table VII. Nominal data handling

Nominal data handling	SDEE Technique	Studies
Dummy variables	Regression	S12, S16, S20, S27
Equality distance	CBR	S9, S20
Dataset segmentation	Regression	S25, S27
interaction	Regression	S27
hierarchical linear model	Regression	S27

Table VIII. Ordinal data handling

Ordinal data handling	SDEE Technique	Studies
Interval scale	Regression / CBR	S12, S20
Grow's formula	CBR	S9
Conversion to numerical values	Regression	S1

Table IX. Other uses of categorical data

Use of categorical data	SDEE Technique	Studies
Adjustment using MT	CBR	S7
Adjustment using ANN	CBR	S19
Weighting using PSO	CBR	S10
Appropriateness of attributes using CORR	CBR	S6
Dataset appropriateness using Mantel's correlation (dataset partitioning based on nominal data)	CBR	S17

*G. Types of used categorical data (MQ7)*

Fig. 4 shows the types of categorical data used in the selected papers. As can be seen, 59% (16 out of 27) of the selected studies dealt with both nominal and ordinal data, 7% (2 out of 27) dealt with only nominal data and 4% (1 out of 27) were concerned with ordinal data. Among the selected studies, 30% (8 out of 27) did not specify the exact

categorical data type that is handled in the paper. However, based on our knowledge and the datasets used in the experiments, we concluded that most of these papers dealt with both nominal and ordinal data types.

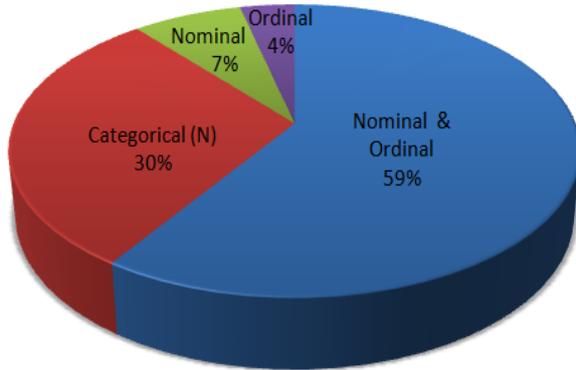


Fig. 4 Types of used categorical data

#### H. Datasets used (MQ8)

Several datasets were used in the selected papers to investigate the use of categorical data in software effort estimation. Table X shows the datasets used for validation as well as the number of studies in which they were used and the percentage of categorical data. The min, max, and mean columns show the minimum value, the maximum value and the mean value respectively of the percentage of categorical data used in the selected papers to conduct experiments. Note that, different studies may opt for different categorical features to conduct experiments. Therefore, the percentage of categorical data is not the same for all studies. Note also that, there were some studies for which it was not possible to extract the percentage of categorical features used in the experiments. As can be seen from table X, 21 datasets were used in the selected papers. Among them ISBSG, COCOMO, Desharnais, Kemerer, Albrecht and Maxwell are the most used datasets. In terms of categorical data percentage, COCOMO (93.52%) was the dataset with the highest mean percentage followed by Maxwell (88.83%) and Laturi (80.00%).

Even if ISBSG is the most used dataset and contains numerous categorical features, the mean percentage of the categorical data used in the selected papers was 49.13%. This is due to the fact that some studies used few categorical features to conduct experiments. Also, there was 1 study [29] that used only the numerical features of ISBSG. This study was included in our mapping study since the technique described in the paper may be applied on both numerical and categorical data. It is worth noting that, some papers employed datasets with numerical and mixed data to show the efficiency of their techniques to deal with both data types.

Table X. Datasets used in the selected papers

Dataset	# of studies	Percentage of categorical data		
		Min	Max	Mean
ISBSG	19	00.00	81.82	49.13
COCOMO	11	88.89	94.74	93.52
Desharnais	9	10.00	25.00	13.10
Kemerer	6	33.00	40.00	36.58
Albrecht	6	00.00	16.67	8.33
Maxwell	5	80.00	95.65	88.83
NASA93	3	60.00	94.44	77.22
Telecom	2	N	N	N
USP05-FT	2	52.94	63.64	58.29
USP05-RQ	2	52.94	63.64	58.29
China	1	00.00	00.00	00.00
DPS	1	00.00	00.00	00.00
CF	1	00.00	00.00	00.00
STTF	1	15.62	15.62	15.62
Laturi	1	80.00	80.00	80.00
Leung02	1	00.00	00.00	00.00
Mends03	1	00.00	00.00	00.00
Atkinson	1	N	N	N
Finnish	1	N	N	N
Mermaid	1	N	N	N
Real-time 1	1	100.00	100.00	100.00

N: Not given in the paper

#### IV. IMPLICATION FOR RESEARCH AND PRACTICE

This study aims at presenting an overview of how categorical data are handled in SDEE. Based on the finding of our SMS, some recommendations to SDEE researchers and practitioners are provided. Dealing with categorical data is an important issue in SDEE especially at the early stages of the software life cycle where most of the existing attributes are more categorical than numerical. This study found that, the publication of SDEE papers with focus on categorical data is characterized by discontinuity. This implies that the use of categorical data in SDEE needs to be more investigated.

No case study was identified in the selected papers. Therefore, it is suggested to the researchers to cooperate with practitioners in order to explore the use of categorical data in industry to yield estimates. We also recommend for researchers to develop tools that enable software effort estimation using a mixture of numerical and categorical data to encourage the use of categorical data by practitioners and researchers.

This study found that CBR, regression and classification and regression trees are the techniques that investigate the most the use of categorical data in SDEE. It is therefore recommended to conduct further research works using other SDEE techniques. Researchers are also encouraged to develop new techniques to handle categorical data instead of

using traditional ones. Furthermore, previous studies revealed that ensemble techniques yield better results than single techniques [26], [45]-[47]. However, all selected papers used single SDEE techniques. No ensemble SDEE technique dealing with categorical data was identified. This implies that researchers should give more attention to the use of categorical data in ensemble techniques to investigate their impact on improving the estimation accuracy of their techniques.

#### V. CONCLUSION AND FUTURE WORK

In this paper, a systematic mapping study was carried out in order to identify and summarize the existing works on SDEE dealing with categorical data. A total of 27 relevant studies were identified and classified according to research approach, contribution type, SDEE technique, Technique used to handle categorical data, types of categorical data and datasets used. Research sources and publication trends were also identified and analyzed. Our findings are summarized as follows.

**(MQ1):** Dealing with categorical data has not been sufficiently investigated in SDEE. Besides, Journals were the most targeted publication channels followed by conferences.

**(MQ2):** The publication of SDEE papers with focus on categorical data is characterized by discontinuity. Dealing with categorical data in SDEE has gained research interest in the period 2008-2013.

**(MQ3):** Solution proposal and history-based evaluation were the two main research approaches used in the selected papers.

**(MQ4):** Most of the selected papers focus on developing new techniques especially to improve existing approaches.

**(MQ5):** Case based-reasoning, regression, fuzzy logic, and classification and regression trees were the techniques that investigate the most the use of categorical data in SDEE.

**(MQ6):** Euclidean distance, fuzzy logic, and fuzzy clustering techniques were the most used techniques to handle categorical data using CBR. Using regression, most papers employed ANOVA and combination of categories.

**(MQ7):** Most of the selected studies dealt with both nominal and ordinal data.

**(MQ8):** ISBSG, COCOMO, Desharnais, Kemerer, Albrecht and Maxwell were the most used datasets.

For future work, we will carry out a systematic literature review to analyze the use of categorical data in SDEE by taking into account the finding of this SMS.

#### REFERENCES

- [1] M. Azzeh, D. Neagu, and P. Cowling, "Software effort estimation based on weighted fuzzy grey relational analysis", in *Proc. 5th International Workshop on Predictive Models in Software Engineering*, Vancouver, BC, Canada, 2009. <https://doi.org/10.1145/1540438.1540450>. S9\*
- [2] I.F. de Barcelos Tronto, J.D. S. da Silva, and N. Sant'Anna, "An investigation of artificial neural networks based prediction systems in software project management", *The Journal of Systems and Software*, vol. 81, pp. 356–367, 2008. <https://doi.org/10.1016/j.jss.2007.05.011>. S26\*
- [3] B.W. Boehm, "Software cost estimation with COCOMOII", NJ: Prentice-Hall, 2000.
- [4] E. Mendes, "The use of Bayesian networks for Web effort estimation: further investigation", in *Proc. 8th Int Conf on Web Engineering*, New York, 2008, pp. 203–216. <https://doi.org/10.1109/ICWE.2008.16>.
- [5] J. Wen, S. Li, Z. Lin, Y. Huc, and C. Huang, "Systematic literature review of machine learning based software development effort estimation models", *Information and Software Technology*, vol. 54, no. 1, pp. 41–59, 2012. <https://doi.org/10.1016/j.infsof.2011.09.002>.
- [6] S.-J. Huang, N.-H. Chiu, and L.-W. Chen, "Integration of the grey relational analysis with genetic algorithm for software effort estimation", *European Journal of Operational Research*, vol. 188, no. 3, pp. 898–909, 2008. <https://doi.org/10.1016/j.ejor.2007.07.002>.
- [7] K.V. Kumar, V. Ravi, M. Carr, and N.R. Kiran, "Software development cost estimation using wavelet neural networks", *Journal of Systems and Software*, vol. 81, pp.1853–1867, 2008. <https://doi.org/10.1016/j.jss.2007.12.793>.
- [8] M.O. Elish, "Improved estimation of software project effort using multiple additive regression trees", *Expert Systems with Applications*, vol. 36, no. 7, pp. 10774–10778, 2009. <https://doi.org/10.1016/j.eswa.2009.02.013>.
- [9] M. Shepperd and C. Schofield, "Estimating software project effort using analogies", *IEEE Transactions on Software Engineering*, vol. 23, no. 12, pp. 736–743, 1997. <https://doi.org/10.1109/32.637387>. S24\*
- [10] M.A. Ahmed and Z. Muzaffar, "Handling imprecision and uncertainty in software development effort prediction: a type-2 fuzzy logic based framework", *Information and Software Technology*, vol. 51, no. 3, pp. 640–654, 2009. <https://doi.org/10.1016/j.infsof.2008.09.004>. S2\*
- [11] R.T. Hughes, "Expert judgment as an estimating method", *Information and Software Technology*, vol. 38, pp. 67–75, 1996. [https://doi.org/10.1016/0950-5849\(95\)01045-9](https://doi.org/10.1016/0950-5849(95)01045-9).
- [12] A. Idri, T. Khoshgoftar, and A. Abran, "Investigating soft computing in case-based reasoning for software cost estimation", *Inter. Jour. of Eng. Int. Sys. for Ele. Eng. and Com.*, vol 10, no. 3, pp. 147-157, 2002.
- [13] A. Idri, A. Abran, and L. Kjiri, "COCOMO Cost Model Using Fuzzy Logic", in *Proc. 7th International conference on Fuzzy Theory and technology*, Atlantic, New Jersey, 2000, pp. 1–4.
- [14] B. Boehm, "Software engineering economics", *IEEE Transactions on Software Engineering*, vol. 10 pp. 4–21, 1984. <https://doi.org/10.1109/TSE.1984.5010193>.
- [15] ISBSG, International Software Benchmark and Standard Group, [www.isbsg.org](http://www.isbsg.org).
- [16] A. Idri, A. Abran, and T. Khoshgoftar, "Fuzzy Analogy: A New Approach for Software Effort Estimation", in *Proc. 11th International Workshop in Software Measurements*, Canada, 2001, pp. 93-101.
- [17] F.A. Amazal, A. Idri, and A. Abran, "Improving Fuzzy Analogy Based Software Development Effort Estimation", in *Proc. 21st Asia-Pacific Software Engineering Conference*, Jeju, South Korea, 1-4 Dec, 2014. <https://doi.org/10.1109/APSEC.2014.46>. S3\*
- [18] L. Angelis, I. Stamelos, and M. Morisio, "Building a Software Cost Estimation Model Based on Categorical Data", in *Proc. 7th International Software Metrics Symposium*, London, UK, 2001, pp. 4–15. <https://doi.org/10.1109/METRIC.2001.915511>. S4\*
- [19] M. Azzeh, D. Neagu, P. Cowling, "Fuzzy grey relational analysis for software effort estimation", *Empirical Software Engineering*, vol. 15, no. 1, pp 60–90, 2010. <https://doi.org/10.1007/s10664-009-9113-0>. S8\*
- [20] A. Idri, F.A. Amazal, and A. Abran, "Accuracy Comparison of Analogy-Based Software Development Effort Estimation Techniques", *International Journal of Intelligent Systems*, vol. 31, no. 2, pp. 128-152, February 2016. <https://doi.org/10.1002/int.21748>. S15\*
- [21] J. Li, G. Ruhe, A. Al-Emran, and M. Richter, "A flexible method for software effort estimation by analogy", *Empirical Software Engineering*, vol. 12, pp. 65–106, 2007. <https://doi.org/10.1007/s10664-006-7552-4>. S18\*

- [22] B. Kitchenham, D. Budgen, and O.P. Brereton, "The value of mapping studies – A participant-observer case study", in Proc. 14<sup>th</sup> International Conference on Evaluation and Assessment in Software Engineering, Keele University, UK, 2010, pp. 1–9.
- [23] B. Kitchenham, O.P. Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering – A systematic literature review", *Information and Software Technology*, vol.51, pp. 7–15, 2009. <https://doi.org/10.1016/j.infsof.2008.09.009>.
- [24] B. Kitchenham, S. Charters, "Guidelines for Performing Systematic Literature Reviews in Software Engineering", Tech. Rep. EBSE-2007-01, Keele University and University of Durham, 2007.
- [25] A. Idri, F.A. Amazal, and A. Abran, "Analogy-based software development effort estimation: A systematic mapping and review", *Information and Software Technology*, vol. 58, pp.206–230, 2015. <https://doi.org/10.1016/j.infsof.2014.07.013>.
- [26] A. Idri, M. Hosni, and A. Abran, "Systematic Mapping Study of Ensemble Effort Estimation", in Proc. 11th International Conference on Evaluation of Novel Software Approaches to Software Engineering, 2016, pp. 132–139. <https://doi.org/10.5220/0005822701320139>.
- [27] M. Azzeh, "Dataset Quality Assessment: An extension for analogy based effort estimation". *International Journal of Computer Science & Engineering Survey (IJCSSES)*, vol.4, no.1, 2013. S6\*
- [28] M. Azzeh, "Model tree based adaption strategy for software effort estimation by analogy", in Proc. of the 11th International Conference on Computer and Information Technology, Pafos, Cyprus, 2011. <https://doi.org/10.1109/CIT.2011.48>. S7\*
- [29] V. K. Bardsiri, D. N. A. Jawawi, S. Z. M. Hashim, and E. Khatibi, "A PSO-based model to increase the accuracy of software development effort estimation", *Software Quality Journal*, vol. 21, no. 3, pp. 501–526, 2013. <https://doi.org/10.1007/s11219-012-9183-x>. S10\*
- [30] L. C. Briand, K. El Emam, D. Surmann, I. Wiecezorek, and K. D. Maxwell, "An Assessment and Comparison of Common Software Cost Estimation Modeling Techniques", in Proc. of the International Conference on Software Engineering (ICSE), Los Angeles, CA, USA, May 1999. <https://doi.org/10.1145/302405.302647>. S12\*
- [31] R. Jeffery, M. Ruhe, and I. Wiecezorek, "Using Public Domain Metrics to Estimate Software Development Effort", in Proc. 7th International Symposium on Software Metrics, April 04 - 06, 2001. <https://doi.org/10.1109/METRIC.2001.915512>. S16\*
- [32] J. W. Keung, B. Kitchenham, and D. R. Jeffery, "Analogy-X: Providing Statistical Inference to Analogy-Based Software Cost Estimation", *IEEE Transactions on Software Engineering*, vol. 34, no. 4, July/August 2008. <https://doi.org/10.1109/TSE.2008.34>. S17\*
- [33] Y. F. Li, M. Xie, and T. N. Goh, "A study of the non-linear adjustment for analogy based software cost estimation", *Empirical Software Engineering*, vol. 14, no. 6, pp. 603–643, December 2009. <https://doi.org/10.1007/s10664-008-9104-6>. S19\*
- [34] L. Haitao, W. Ru-xiang, and J. Guo-ping, "Similarity measurement for data with high-dimensional and mixed feature values through fuzzy clustering", in Proc. International Conference on Computer Science and Automation Engineering (CSAE), 2012. <https://doi.org/10.1109/CSAE.2012.6273028>. S13\*
- [35] S.-J. Huang, C.-Y. Lin, and N.-H. Chiu, "Fuzzy Decision Tree Approach for Embedding Risk Assessment Information into Software Cost Estimation Model", *Journal of Information Science and Engineering*, vol. 22, pp. 297–313, 2006. S14\*
- [36] N. Mittas, and L. Angelis, "LSEbA: least squares regression and estimation by analogy in a semi-parametric model for software cost estimation", *Empirical Software Engineering*, vol.15, pp. 523–555, 2010. <https://doi.org/10.1007/s10664-010-9128-6>. S20\*
- [37] P. Sentas, L. Angelis, I. Stamelos, and G. Bleris, "Software productivity and effort prediction with ordinal regression", *Information and Software Technology*, vol. 47, pp. 17–29, 2005. <https://doi.org/10.1016/j.infsof.2004.05.001>. S22\*
- [38] M. Azzeh, "Software Effort Estimation Based on Optimized Model Tree", in Proc. 7th International Conference on Predictive Models in Software Engineering, Banff, Alberta, Canada, September 20-21, 2011. S5\*
- [39] S. Bibi, I. Stamelos, and L. Angelis, "Combining probabilistic models for explanatory productivity estimation", *Information and Software Technology*, vol. 50, pp. 656–669, 2008. <https://doi.org/10.1016/j.infsof.2007.06.004>. S11\*
- [40] E. Papatheocharous, and A. S. Andreou, "Classification and Prediction of Software Cost through Fuzzy Decision Trees". in Proc. International Conference on Enterprise Information Systems, 2009, pp. 234-247. [https://doi.org/10.1007/978-3-642-01347-8\\_20](https://doi.org/10.1007/978-3-642-01347-8_20). S21\*
- [41] M. Tsunoda, S. Amasaki, and A. Monden, "Handling categorical variables in effort estimation", in Proc. 2012 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement, 20-21 Sept. 2012. <https://doi.org/10.1145/2372251.2372267>. S27\*
- [42] P. Silhavy, R. Silhavy, and Z. Prokopova, "Categorical Variable Segmentation Model for Software Development Effort Estimation", *IEEE Access*, vol. 7, pp. 9618 - 9626, 11 January 2019. <https://doi.org/10.1109/ACCESS.2019.2891878>. S25\*
- [43] R. Abdulkalykov, I. Hussain, M. Kassab, and O. Ormandjieva, "Quantifying the Impact of Different Non-functional Requirements and Problem Domains on Software Effort Estimation", in Proc. Ninth International Conference on Software Engineering Research, Management and Applications, Baltimore, MD, USA, 2011. <https://doi.org/10.1109/SERA.2011.45>. S1\*
- [44] Y. Shan, R. I. McKay, C.J. Lokan, and D.L. Essam, "Software Project Effort Estimation Using Genetic Programming", in Proc. International Conference on Communications, Circuits and Systems and West Sino Expositions (ICCCAS), Chengdu, China, 29 June-1 July 2002. <https://doi.org/10.1109/ICCCAS.2002.1178979>. S23\*
- [45] A. Idri, M. Hosni, and A. Abran, "Systematic Literature Review of Ensemble Effort Estimation", *Journal of Systems and Software*, vol. 118, pp. 151–175, 2016. <https://doi.org/10.1016/j.jss.2016.05.016>.
- [46] M. Hosni, A. Idri, and A. Abran, "Investigating Heterogeneous Ensembles with Filter Feature Selection for Software Effort Estimation", in Proc. 27th International Workshop on Software Measurement and 12th International Conference on Software Process and Product Measurement, ACM, New York, NY, USA, 2017: pp. 207–220. <https://doi.org/10.1145/3143434.3143456>.
- [47] M. Azzeh, A.B. Nassif, and L.L. Minku, "An empirical evaluation of ensemble adjustment methods for analogy-based effort estimation", *Journal of Systems and Software*, vol. 103, pp. 36–52, 2015. <https://doi.org/10.1016/j.jss.2015.01.028>.

# Big Data Platform for Smart Grids Power Consumption Anomaly Detection

Peter Lipčák<sup>2</sup>, Martin Macak<sup>1,2</sup> and Bruno Rossi<sup>1,2</sup>

<sup>1</sup>*Institute of Computer Science, Masaryk University*

<sup>2</sup>*Faculty of Informatics, Masaryk University*

Brno, Czech Republic

Email: {plipcak,macak,rossi}@mail.muni.cz

**Abstract**—Big data processing in the Smart Grid context has many large-scale applications that require real-time data analysis (e.g., intrusion and data injection attacks detection, electric device health monitoring). In this paper, we present a big data platform for anomaly detection of power consumption data. The platform is based on an ingestion layer with data densification options, Apache Flink as part of the speed layer and HDFS/KairosDB as data storage layers. We showcase the application of the platform to a scenario of power consumption anomaly detection, benchmarking different alternative frameworks used at the speed layer level (Flink, Storm, Spark).

## I. INTRODUCTION

**B**IG data architectures are designed to handle ingestion, processing and analysis of data that possesses the five V's properties: Volume, Velocity, Value, Variety, Veracity [1], [2]. In the context of Smart Grids (SG), utilities have to deal with an increasing volume of data leading to typical big data problems. According to Zhang et al. [3], the five V's in the SG domain are represented by several needs: to analyze large amounts of data in real-time, like from smart meter readings (Volume), to deal with quick generation of records (Velocity), and diversity of data structures (Variety), with multiplicity of use case, such as anomaly detection or load balancing that bring value to the customers (Value), and inherent problematic of data in terms of possible measurement errors (Veracity). To exemplify, with a sampling rate of 15 minutes, a sample of 1 Million Smart Meter devices installed results in around 3 Petabytes of data in one year (3000TB, ~35Billion records at a size of 5KB each record) [4].

There are a plethora of use cases for the application of big data analysis in the context of SGs [5], [6], like anomaly detection methods to detect power consumption anomalous behaviours [7], [8], the analysis of false data injection attacks [9], load forecasting for efficient energy management [10], among others. Such data analysis requirements create needs to define architectures and platforms to support large scale data analysis.

In this paper, we focus on power consumption data anomaly as the application scenario: dealing with the identification of anomalous patterns from energy consumption traces collected from smart meters, that can have several benefits for utilities, such as load optimizations based on determined patterns of energy usage [5], [7] or clustering of customers [11]. The final

goal is the definition and evaluation of a big data platform for power consumption anomaly detection.

We have two main contributions in this paper:

- the provision of a big data platform for power consumption anomaly detection with the main components mapped to the reference architecture proposed by Pääkkönen and Pakkala [12].
- the results of a scenario run with public datasets to assess the applicability of batch-oriented (Apache Spark), stream-oriented (Apache Storm), or hybrid (Apache Flink) frameworks in the speed layer of the platform.

The paper is structured as follows. In Section II, we discuss the background of big data analysis in the context of Smart Grids. In Section III, we discuss big data energy management platforms that can be comparable to our proposal. In Section IV, we propose a platform for big data power anomaly detection with components mapped to the reference architecture in Pääkkönen and Pakkala [12]. In Section V, we propose a power consumption scenario aimed at showcasing the application of the platform and the evaluation at the speed layer level of three frameworks from the Apache Software Foundation (Spark, Flink, Storm). The conclusions are presented in Section VI.

## II. BACKGROUND - BIG DATA ANALYSIS ARCHITECTURES

Big data processing is assuming more and more relevance in many fields of modern society. For energy utilities, the needs to manage energy resources based on the vast amount of information collected from sensors and the ICT infrastructure is nowadays of paramount importance [3].

In the context of big data processing, we can have a first distinction between batch and stream processing. Batch processing is a type of processing executed on large blocks (batches) of data stored over a period of time. These data blocks are appended to highly scalable data stores and periodically analyzed in batches by big data processing frameworks. This approach to data processing is very effective in case of large datasets for use cases that are not time-critical, as the main drawbacks are higher latencies in processing requests [13]. On the other hand, stream processing allows dealing with data in real-time, getting approximate results that can be complemented, if needed, from the analysis of batch processing. Managing streams of data usually implies the capabilities of online learning. Some authors also consider

an intermediate category: micro-batch processing [13], that overcomes some of the issues of batch and stream processing: near real-time performance is granted by considering streams of data in micro-batches sent to the batch processing engine.

There are two popular architectures that were proposed over time, Lambda and Kappa [13], [14]. They were mainly based on the relevance that is given to batch and stream processing. **Lambda architecture** is a processing architecture designed to handle massive amount of data efficiently by taking advantage of both batch and stream-processing methods. Efficiency in this context means high-throughput, fault-tolerance and low latency [13]. The rise of the Lambda architecture is correlated with the growth of data and the speed at which they are being generated, real-time analytics and the drive to mitigate big latencies of map-reduce [15].

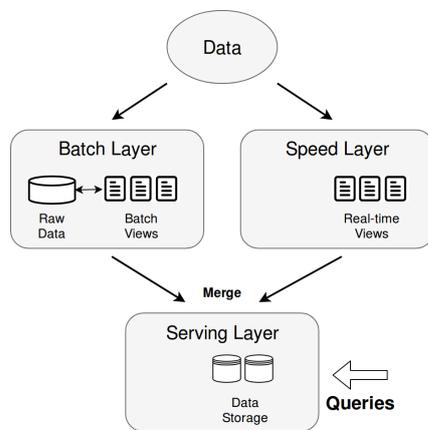


Fig. 1. Lambda Architecture

Generally, a Lambda architecture consists of three distinct layers (Fig. 1): a batch layer (batch processing), a speed layer (stream processing) and a serving layer (data storage). The batch layer is responsible for bringing comprehensive and accurate views of batch data while simultaneously, the speed layer provides near-real-time data views. Stream processing can take advantage of batch views and may be joined before presentation. Data streams entering the system are dual fed into both batch and speed layer.

The batch layer stores raw data as it arrives and computes the batch views in intervals. When the data gets stored in the data store using different data storage systems, the batch layer processes the data using one of the big data processing frameworks that implement the map-reduce programming paradigm. Popular frameworks that support batch processing are Apache Hadoop, Apache Spark, and Apache Flink.

The speed layer processes data streams in real-time with the focus on minimal latencies. Usually latencies vary from milliseconds to several seconds. This layer often takes advantage of pre-computed batch views. Popular frameworks that support stream processing are Apache Flink, Apache Storm, Apache Spark, and Apache Samza.

The serving layer aggregates the outputs from batch and speed layers, storing the data in a datastore. As storage, highly

scalable and distributed data lakes are often used. Among popular big data datastores belong Apache Cassandra, Apache HBase, Hadoop Distributed File System, OpenTSDB, and KairosDB.

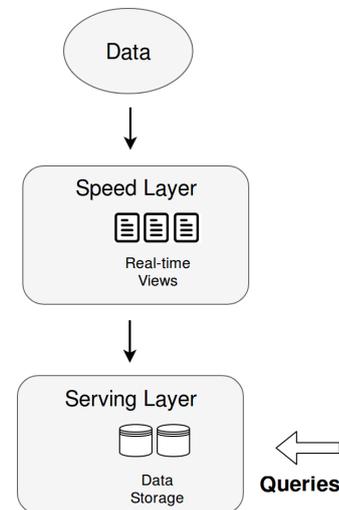


Fig. 2. Kappa Architecture

**Kappa architecture** is an alternative to the Lambda architecture proposed to overcome some of the limitations, like maintaining two code bases for batch and speed layers and the general complexity of the platform [16]. The Kappa architecture consists of two distinct layers (Fig. 2): a speed layer and a serving layer. The speed layer processes streams of data in the same way as Lambda architecture. The only main difference is that when the code changes, data needs to be reprocessed again. This is because parts of the Kappa architecture act as an online learner.

Big data analysis frameworks often do not support both batch and stream processing, thus a hybrid combination of frameworks is the choice, e.g., Apache Hadoop and Apache Storm can be used to fully support a Lambda architecture [17].

In the context of SGs, we can find some examples of the main constituting layers of both architectures. A batch layer in Spark, distributed in-memory computing framework, was used to pre-compute a statistical-based model using linear regression for anomaly predictions of energy consumption data [7]. A stream layer was used for the detection of defective smart meters, solution implemented using the Flink stream processing engine [18]. As a serving layer, KairosDB, time-series database built on top of Apache Cassandra, could handle the workload of a large city with around six million smart meters. During this experiment, KairosDB was installed in a cluster of 24 nodes [19].

The next section will discuss more in detail about big data energy management platforms that have been proposed so far.

### III. RELATED WORKS—BIG DATA ENERGY MANAGEMENT PLATFORMS

We have discovered several proposed big data architectures in the SG domain that we mapped to either Lambda or Kappa based on the available information. We found that the majority of the architectures are cloud-oriented and that several energy management architectures do not specify the applicability to the big data context. Therefore, they might use different architectural structure than pure Lambda or Kappa.

#### A. Big Data energy management architectures

Mayilvaganan and Sabitha [20] proposed a SG architecture which uses HDFS and Cassandra to store historical data for the prediction of energy supply and demand. In this work, only MapReduce processing was used.

Munshi and Mohamed [21] presented a SG big data ecosystem based on the Lambda architecture. Smart meter data are being ingested to a cloud with Flume. For the batch layer of the Lambda architecture Hadoop is used, and Spark for the speed layer. Authors also performed data mining and visualization applications on top of this ecosystem with real data.

Liu and Nielsen [22] designed a smart meter analytic system. The architecture is divided into data ingestion, processing, and analytics layer. It can process both batch and stream data. In the processing module, they list several tools which can be used, like Spark, Hive, or Python. In the end, the data are sent to the analytics layer, which contains a PostgreSQL database, analytics libraries, and applications for users. This architecture can be viewed as Lambda-based because usage of Hive can be considered as a batch layer and the usage of Spark as a speed layer. The analytics layer of this architecture can be mapped to a serving layer.

Fernández et al. [?] proposed an architecture that improves energy efficiency management in a smart home. It consists of four modules: data collection, data storage, data visualization, and a machine learning module. It is designed to work with both batch and real-time processing. The data storage module consists of three blocks: acquisition, real-time, and batch block. From those processing blocks, the data are available to blocks that can be mapped to a service layer of a Lambda architecture. Therefore this architecture is also considered as Lambda-based.

Balac et al. [23] proposed an architecture for real-time predictions in energy management. The data streams are collected to the server, which provides management functionality like dashboards, alerting, and basic reporting. From this server, data are transferred to their high-performance file system where the real-time analysis is performed. The analysis results are then passed back to the server but are also archived in the cloud storage for some later batch processing task. Based on this description, this architecture can be viewed more as Kappa-based.

Al-Ali et al. [24] proposed a system for energy management in a smart home. They specify both hardware and software architecture. The software architecture contains three modules: data acquisition, a middleware, and a client application

module. In this architecture, data are stored and then used by several services. Therefore it does not represent neither a Lambda nor a Kappa architecture.

Daki et al. [25] presented an architecture which is composed of five parts: data sources, integration, storage, analysis, and visualization that can be used for the analysis of customer data. They provide a set of technologies which might be used and can be considered as either a Lambda or a Kappa architecture, depending on the use cases.

#### B. Other energy management architectures

Yang et al. [26] proposed an energy management system that uses a service-oriented architecture in the cloud. For storage, they use distributed software as MySQL Cluster and HDFS. However, authors do not specifically mention big data. Ali et al. [27] proposed a computing grid based framework for the analysis of system reliability and security. The architecture consists of three layers: application, grid middleware, and resource layer, with focus on high-performance computing. Rajeev and Ashok [28] presented a cloud computing architecture for power management of microgrids, consisting of four modules: infrastructure, monitoring, power management, and a cloud service module.

### IV. PROPOSED PLATFORM

In this section, we propose our big data platform for smart meters power consumption anomaly detection, based on our previous research ([29], [8], [30], [31]). The goal of such platform is to process large amounts of data from smart meters and weather information sources to detect anomalous behaviours from the side of customers. Such analysis can allow to further create customer profiles that can be used to cluster users according to their power consumption behaviours [11]. The five V's in this area derive from several aspects, namely volume: the large amount of information traces generated by smart meters multiplied by the number of users [4], velocity: the needs for real-time analysis of such traces that are constantly generated [4], variety: multiple data sources involved, either structured or unstructured, mainly about power consumption and weather data [3], value: the added value that analyses can have for utilities, that can create customer profiles to optimize power production and balancing of the whole grid [11], and veracity: the many issues that measurements errors might pose, such as corrupted or missing data from smart meters [32].

We map our proposed architecture to the reference architecture by Pääkkönen and Pakkala [12]. The reference architecture for big data systems is technology independent and is based on the analysis of published implementation architectures of several big data use cases. We mapped all the functionalities, data stores and data flows contained within our platform proposal to the reference architecture diagram to allow for easier comparison with other platforms (Fig. 3):

**Data sources.** Our platform supports two possible data sources. First, a stream of semi-structured data is being collected from smart meters datasets. These can either be live data

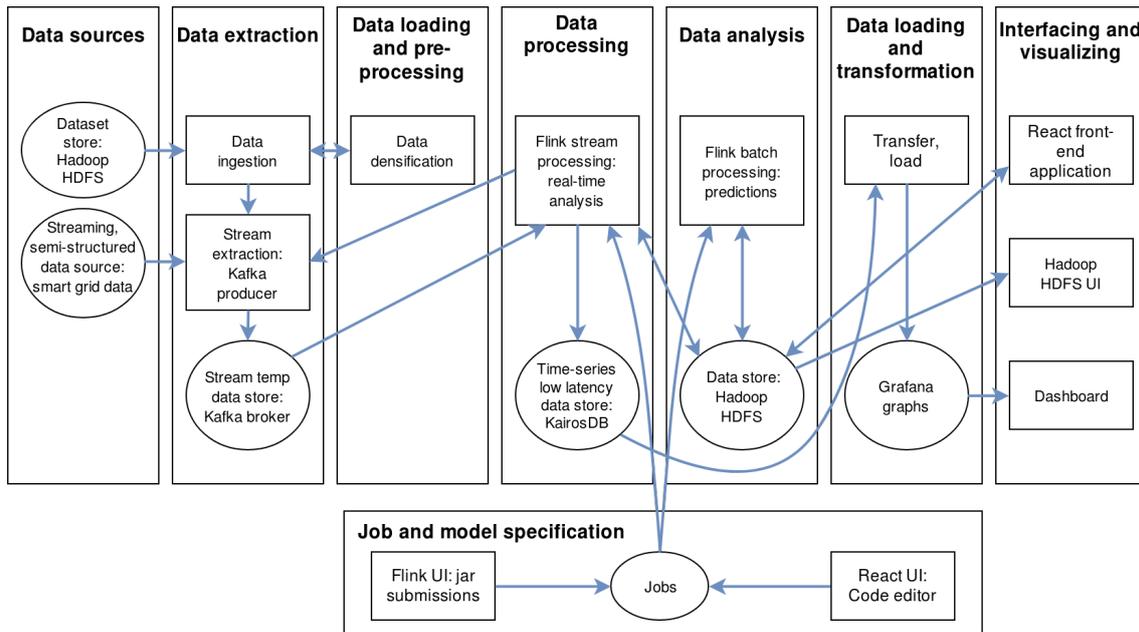


Fig. 3. Architecture Mapping to Pääkkönen and Pakkala [12] Reference Architecture

from smart meters or private/public available datasets. Another way for data to enter the platform is reading data available in the Hadoop Distributed File System (HDFS).

**Data extraction.** A Kafka producer corresponds to the stream extraction functionality and a Kafka broker serves as temporary data store. An ingestion manager extracts data from HDFS and sends data to the Kafka broker.

**Data loading and pre-processing.** Data extracted from HDFS using the ingestion manager can be further densified. Densification can increase the itemset for smaller datasets (e.g., for benchmarking reasons). Currently, two densification methods are available in the platform – multiplication (replicating  $n$  times the dataset) and interpolation (constructing new data points based on interpolating previous intervals  $n$  times). However, more advanced methods can be implemented, such as regression-based and probability-based methods [33].

**Data processing.** Flink's stream-processing jobs read incoming data streams from Kafka. These jobs can also access HDFS to read pre-computed models or datasets and merge them before producing results. The output of stream-processing jobs can be sent back to Kafka, HDFS or to the time-series datastore KairosDB.

**Data analysis.** Flink's batch-processing jobs can read data from HDFS and perform data analytics. The output can be further stored in HDFS.

**Data loading and transformation.** Results of stream-processing jobs stored in KairosDB can be loaded into a Grafana server. In this context, Grafana server serves as a temporary data store until the graphs are generated.

**Data storage.** The following data storage technologies (temporary or persistent) are supported by the platform: Kafka broker, HDFS, KairosDB and Grafana.

**Interfacing and visualization.** There are several user interfaces that allow interacting with the platform. HDFS provides an UI with information regarding storage and the file system. Flink UI gathers statistics about running jobs, e.g. records processed per second by an operator. A dashboard component displays graphs produced by Grafana. A React front-end application allows users to upload or delete datasets and shows datasets previews. It is also possible to select the dataset for ingestion, which invokes the ingestion Manager to read the dataset from HDFS and ingests the data into Kafka.

**Job and model specification.** Submitting stream and batch processing jobs can be done either by uploading a JAR file with all dependencies using the Flink UI or by submitting the code using an ad-hoc code-editor contained within a React front-end application. The JAR file has to include the source code for the processing job (e.g., an anomaly detection algorithm like in [7]), all the dependencies the job requires and the path to the entry Java file to be executed.

Discussing the platform's architecture (Fig. 4), data itemsets generated by smart meter devices are forwarded to the platform using a publish/subscribe messaging system. Kafka can be used as a data source and a data sink for Flink's jobs and the Kafka Connector provides access to event streams without manual implementation needs — making it a good choice for the platform. Each of the technologies is capable of running in clusters, providing scalability and fault-tolerance. Each technology can be scaled independently based on the use cases, making the platform flexible in terms of configuration.

Apart from the integration of existing big data tools, we implemented three separate applications to support the needs of smart meters data analysis (Fig. 4). These applications are: i) an ingestion manager, responsible for densification and

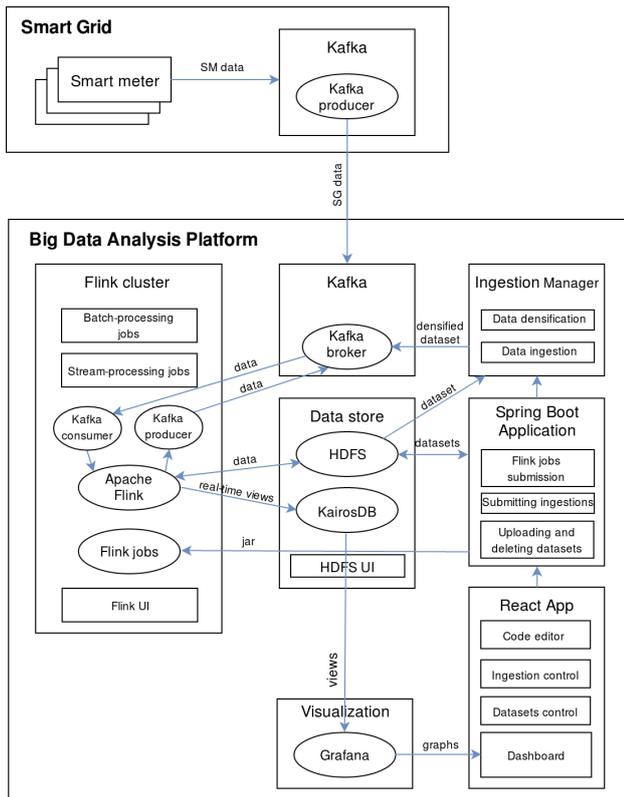


Fig. 4. Platform Architecture

ingestion of datasets into Kafka, ii) a Spring Boot Application, i.e. a back-end implementation with a RESTful API which enables execution of ingestions, Apache Flink job submissions, and uploading and deleting datasets to/from HDFS, iii) a React front-end application that allows users to submit ingestions, upload and delete datasets. Apache Flink provides both batch and stream processing of the data, thus a Lambda architecture is fully supported, as discussed in the background section. As the main data storage, HDFS was chosen and KairosDB is used to provide real-time views on the data. Grafana can be integrated with KairosDB and produce real-time graphs of data. Generated graphs can be further shown in the dashboard of the application.

### V. POWER CONSUMPTION ANOMALY DETECTION SCENARIO

We showcase the use of the platform in the context of power consumption anomaly detection. Studying unusual consumption behaviors of customers and discovering unexpected patterns is an important topic related to the use of smart metering devices in the smart grids domain, as discussed in the previous section and in related research ([7], [8]).

In this context, we propose a scenario to showcase the platform and to look into the performance of three different frameworks for the streaming part: batch-based (Spark), stream-based (Storm), and hybrid (Flink). As the speed layer

is a key part of the performance of the platform, the selection of the best framework is an important decision.

#### A. Compared Frameworks

**Apache Spark.** Created in 2009, is a general purpose processing engine suitable for a wide range of use cases. There are four libraries built on top of Spark processing engine: Spark SQL for SQL language support, Spark Streaming for stream processing support, MLlib for machine learning algorithms, GraphX for graph computations. Languages supported by Spark include: Java, Python, Scala, and R.

Spark applications consist of a driver program that runs the main function and executes various operations in parallel. The main abstraction that Spark provides is a resilient distributed dataset (RDD), which is a collection of elements partitioned across the cluster nodes. Operations such as map or filter executed on RDDs are executed in parallel. Spark Streaming discretizes the streaming data into micro-batches, then latency-optimized Spark engine runs short tasks to process the batches and outputs the results. Each batch of data is an RDD [34].

**Apache Flink.** Created in 2009, is a framework and distributed processing engine for computations on both batch and streams of data. Stream processing is supported natively and provides excellent performance with very low latencies. It also provides a machine learning library called FlinkML as well as a graph computation library Gelly. Supported programming languages are Java, Scala, Python, and SQL.

Flink provides different levels of abstraction that can be used to develop batch or stream processing applications. Abstractions from low-level to high-level respectively are as follows: Stateful Streaming Processing, DataStream / DataSet API, Table API and SQL. In practice, most applications would not need the lowest-level abstraction, Stateful Streaming Processing, but would rather program against the Core APIs like the DataStream API (bounded/unbounded streams) and the DataSet API (bounded datasets). These APIs provide very similar operations as Spark's RDDs such as map, filter, aggregation and other transformations [35].

**Apache Storm.** Created in 2011, is a distributed real-time processing engine. Storm is a pure stream processing framework without batch processing ability. Storm provides great throughput with very low latencies. It was designed to be usable with any programming language thanks to its Thrift definition for defining and submitting topologies. Zookeeper is also required to be installed because Storm uses it for cluster coordination.

The overall logic of Storm applications is packaged into a topology. A Storm topology is analogous to a MapReduce job with the difference that Storm applications run forever. A topology is an acyclic directed graph (DAG) composed of spouts and bolts connected with stream groupings. The stream is a core abstraction in a Storm application. A stream is an unbounded sequence of tuples that are generated and processed in parallel. Spouts serve as a source of streams in a topology. Analogically to Flinks DataStream/DataSet or Sparks RDDs operations, bolts are responsible for all the

distributed processing. Bolts can implement operations such as map, filter or aggregation [36].

We could not find any specific benchmarking study of the three frameworks based on smart grids related datasets, but previous studies found contrasting results in other domains.

Karimov et al. [37] found Flink to have more than three times faster throughput than Spark and Storm for aggregations. Joins were more than two times faster for Flink than Spark. Flink outperformed Storm and Spark in six out of seven benchmark categories including throughput and latency. However, Spark was found to perform better than the two other frameworks in case of skewed data, as well to improve the performance more than the other frameworks in presence of more than three nodes.

Wang et al. [38] developed a full benchmarking system to test the performance of Storm, Flink, and Spark. The results of the application of the benchmark showed that Flink is three times faster than Spark and six times faster than Storm in processing advertisement clicks. The final conclusion is that Storm would be the best choice if very low latency is requested, while Spark would be a good option if throughput is a key aspect of the use case. Flink gives a more balanced performance with low latency and high throughput.

Lopez et al. [39] found that Storm and Flink were consistently better than Spark in terms of throughput. Storm had in general better throughput than the other frameworks, while Spark had the worse performance due to the application of micro-batches, as each batch is grouped before processing. However, Spark was found to be more reliable in terms of node failures and recoverability of the functionality. The conclusion is that the lower performance of Spark Streaming might be justified in use cases in which absolute reliability is necessary, considering no messages loss in case of nodes failures.

Chintapalli et al. [40] performed a benchmarking of Spark, Flink, Storm Streaming focusing on latency. Storm performed the best with Flink, both having less than 2 seconds latency at high throughput. The latencies of Spark, on the other hand, were rising with higher throughput, resulting in latencies of over 8 seconds.

### B. Scenario definition

Our scenario contains three implementations of the same algorithm using the three different big data processing frameworks (Spark, Storm, Flink). Implementing the same use case we can easily compare the performance of each framework by measuring the processing time of the same dataset.

**Dataset.** Our first dataset consists of power consumption data collected from apartments in one building with a sampling rate of 15 minutes [41]. The apartment dataset  $(id, timestamp, consumption(kW))$  contains data for 114 single family apartments for the period 2014-2016. The second dataset contains weather data  $(timestamp, temperature, humidity, pressure, windspeed, \dots)$  with a sampling rate of one hour. The size of the apartment dataset is 2.1 GBs and contains 64 million records. For the scenario to better

represent a big data context, we replicated each of these records eight times. The ingestion manager part of the platform with multiplication densification method was used for this purpose. The result size of the testing dataset is 512 million records in CSV format.

**Scenario Setup.** The scenario was run with three server nodes: a Master node and two worker nodes (Fig. 5). Apart from Kafka, each technology runs in a cluster on all three nodes. Kafka was only running in the Master node, as we considered it could serve all other nodes without delays. For Kafka to operate, there is the need to have Apache Zookeeper to manage the cluster. Because Apache Storm uses Zookeeper for cluster management as well, we decided to install Zookeeper in all three nodes. Each of these nodes was configured with an Intel Xeon E3 (2.4GHz, 4 cores), 8GB RAM running on Ubuntu 16.04 LTS 64bit Linux 4.4.0 kernel.

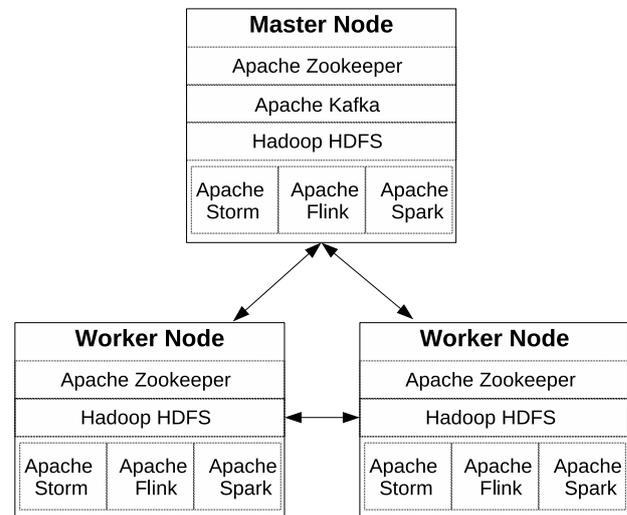


Fig. 5. Nodes involved in the experimental setup

**Anomaly Detection Algorithm Implementation.** We implemented a simple algorithm for finding consumption anomalies using a pre-computed model of consumption predictions (pseudocode, algorithm 1). For the computation of anomaly detection, we implemented a Spark batch processing program in the Java programming language. We did not implement this algorithm using other big data processing frameworks because our main focus was on measuring performance of stream processing. To decide whether a current power consumption item is an anomaly, we analyse power consumptions of three previous days. We take each hour of a day as a season, i.e.,  $t=[0, 23]$ , and use previous three days consumptions at the time  $t$  to compute our predictions. We also take into consideration outside weather because it is highly correlated with power consumption: during winter power consumption is higher due to heating, as well as in summer due to the cooling equipment in function. For each apartment, predictions are made individually because customers have different living habits and power consumption predictions cannot be

generalized to all of them.

```

Input:  $C, T$            ▷ Consumption and Temperature Datasets
Output:  $P$            ▷ Anomaly Detection Model
function CreateAnomalyDetectionModel( $C, T$ ):
     $P \leftarrow \emptyset$            ▷ Initialize Anomaly Detection Model
    foreach  $d \in \text{days}(2014 - 2016)$  do
        foreach  $s \in 0..23$  do
            foreach  $id \in \text{apartmentIds}$  do
                 $C1 \leftarrow \text{MakeAverage}(\text{GetConsumptions}(d - 1, s, id))$ 
                 $XT1 \leftarrow \text{Compute.XT}(\text{GetOutsideTemperature}(d - 1, s))$ 
                 $C2 \leftarrow \text{MakeAverage}(\text{GetConsumptions}(d - 2, s, id))$ 
                 $XT2 \leftarrow \text{Compute.XT}(\text{GetOutsideTemperature}(d - 2, s))$ 
                 $C3 \leftarrow \text{MakeAverage}(\text{GetConsumptions}(d - 3, s, id))$ 
                 $XT3 \leftarrow \text{Compute.XT}(\text{GetOutsideTemperature}(d - 3, s))$ 
                 $\text{Prediction} \leftarrow (C1 * XT1 + C2 * XT2 + C3 * XT3) / 3 + 7$ 
                 $P.\text{insert}(d, s, id, \text{Prediction})$ 
            end
        end
    end
    return  $P$ 

```

**Algorithm 1:** Anomaly Detection Model Pseudocode:  $s$ =season,  $n$ =day,  $C$ =avg power consumption,  $XT$ =outside temp variables

The evaluation of the performance of the frameworks happens at the speed layer. The speed layer of our Lambda implementation is taking advantage of the pre-computed model (algorithm 1), based on which we can detect anomalies of incoming power consumption data. First, we load the anomaly detection model from the datastore and then we compare smart meter readings against this model. If the new consumption value exceeds the value from the model, we consider it as an anomaly (pseudocode, algorithm 2).

We implemented the stream processing part in Java using each framework (Spark, Flink, Storm). Each implementation contains framework specific distributed operations such as map, filter, foreach.

```

Input:  $M, C$            ▷ Anomaly Detection Model and New
                        Consumption
Output:  $A$            ▷ Anomalies
function AnomalyDetection( $M, C$ ):
     $A \leftarrow \emptyset$            ▷ Initialize Anomalies
    foreach  $c \in C$  do
         $P \leftarrow M.\text{get}(c.\text{day}, c.\text{season}, c.\text{id})$ 
        if  $c.\text{value} > P$  then
             $A.\text{insert}(c.\text{day}, c.\text{season}, c.\text{id}, P)$ 
        end
    end
    return  $A$ 

```

**Algorithm 2:** Anomaly Detection Streaming Process

**Process Flow of Anomaly Detection.** The flow of the process of anomaly detection in this scenario is as follows. The power consumption dataset, as well as the weather dataset, were both stored in HDFS. First, Spark loads the datasets into

memory and performs operations in a distributed environment to generate the prediction model using Algorithm 1 (Fig. 6, steps 1-2). The output of Spark is stored back into HDFS (Fig. 6, step 3).

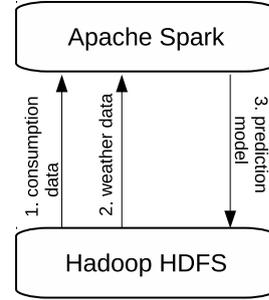


Fig. 6. Anomaly Detection Model Data Flow

**Real-time Anomaly Detection Data Flow.** The data flow of real-time anomaly detection is shown in Fig. 7. The prediction model is first loaded into memory from HDFS (Fig. 7, step 1). After the big data processing framework in use (Spark, Storm, Flink) is initialized and ready, we can start streaming data into Kafka. This is done using the ingestion manager. The ingestion manager reads consumption data from HDFS and sends each record multiple times to Kafka, in this scenario eight times (Fig. 7, steps 2-3) using the multiplication densification method. Big data processing frameworks subscribe to consumption topic and right after new data arrive, they start processing (Fig. 7, step 4). All found anomalies are sent to the Kafka topic "anomalies" (Fig. 7, step 5).

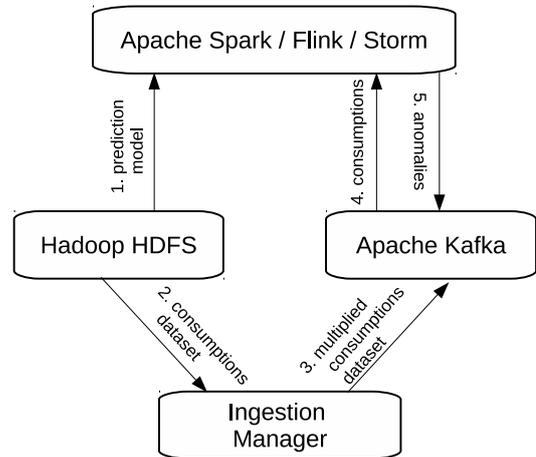


Fig. 7. Real-time Anomaly Detection Data Flow

**C. Results**

We can get several insights about running the platform with each of the three frameworks as the speed layer (we summarize them in Table I).

**Throughput.** Records per seconds processed (Fig. 8) showed that Storm (~378k records per second) and Flink

(~438k records per second) were significantly faster than Spark Streaming (~168k records per second). For the power consumption dataset used in the scenario (512 million records), this means average times of ~20min. (Flink), ~25min. (Storm) and ~50min. (Spark). If throughput is important, Flink seems to give the best results.

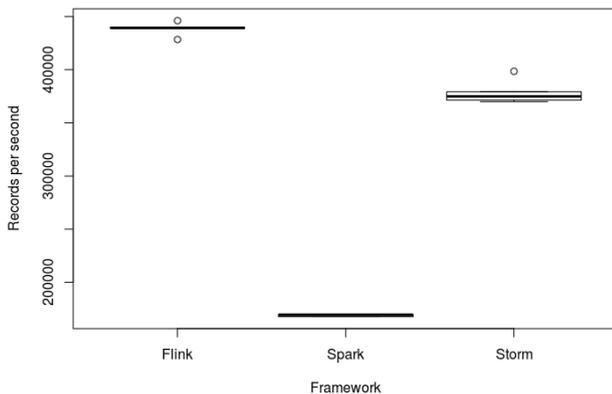


Fig. 8. Throughput. Records per seconds processed (5 runs per framework)

**Latency.** Internally, Spark Streaming receives live data from various sources and divides them into batches (micro-batches), which are then processed by the Spark engine to generate a stream of results. Thus, it is not considered as native streaming, but this way it can also efficiently support processing of big streams of data. However, Spark Streaming strongly depends on the batch intervals which can range from hundreds of milliseconds. During the scenario runs, both Flink and Storm had lower latencies, in terms of milliseconds, while Spark had latencies in terms of seconds. However, the selection depends on how important the batch layer (or micro-batch) is for the specific scenario. For the current scenario using a pre-computed (batch-level) model or an online learning algorithm at the speed layer, both Flink and Storm can be better options than Spark, considering latency.

**Support for batch processing.** In the scenario, the batch layer is only used for a pre-computed model, if updating the model is necessary, Spark natively supports batch processing and provides a very efficient batch processing engine. Batch processing in Flink is dealt with as a special case of stream processing. Apache Flink provides streaming API that can do both bounded and unbounded use cases, but also offers different DataSet API and runtime stack that is faster for batch processing use cases, so it is also possible to process batches very efficiently. Storm does not support batch processing in the current version. In this scenario, the pre-computed model was managed by Spark, so a hybrid combination of frameworks is necessary for a Lambda architecture, if adopting Storm.

**Effort to set-up and configure each framework.** Setting up the cluster nodes for the scenario with the default configuration requires very little time, several minutes to one hour for

TABLE I  
SUMMARY OF THE SPEED LAYER FRAMEWORKS COMPARISON

	Spark	Flink	Storm
Performance	Medium	Very good	Good
Latency	Medium	Very low	Very low
Batch processing support	Yes	Yes	No
Cluster configuration effort	High	Medium	Medium
Scale-up effort	Low	Low	Low
Machine learning support	Very good	Good	Medium

all the three frameworks, Spark, Flink, and Storm. However, fine-tuning each framework brings different considerations. Spark is very flexible and allows to fine-tune many aspects that require a deep knowledge of the framework's architecture. Such fine-tuning can require a large amount of effort. Also Flink requires some effort to tune up the configuration for better performance. This process can take some considerable amount of time, although we found to be simpler in some configuration aspects, as the flexibility of Spark can have drawbacks for the many parameters that can be configured. For Storm there are similar considerations to Flink in terms of configuration, understanding the architecture of the framework is essential for optimization, but also running some tests can give an evaluation of which parameters can give better results.

**Effort required to scale-up the framework.** Each of the three frameworks is relatively easy to scale-up to more nodes. For each framework is a matter of changes in the configuration and propagating the changes to the newly added nodes. The represented scenario can be scaled-up to use more nodes.

**Machine learning libraries and algorithms supported.** While in the scenario we used a simple anomaly detection algorithm, if advanced machine learning operations are necessary, Spark is the framework with the best support. Spark comes with a machine learning library called MLlib that provides common machine learning functionalities and multiple types of machine learning algorithms, such as classification, regression, clustering, etc. All of this is designed to distribute the computing across the cluster. FlinkML is a machine learning library for Flink. It provides algorithms for supervised and unsupervised learning, recommendations and more. The list of supported algorithms is still growing and there is an ongoing work in this area. Storm does not come with any machine learning library, but there is an ongoing work on third-party library called SAMOA, that adds machine learning support to Storm. SAMOA is currently undergoing incubation process in Apache Software Foundation and provides a collection of algorithms for most common data mining and machine learning tasks such as regression, classification, and clustering.

**Threats to Validity.** There are several threats to validity that we need to report. For internal validity, the configuration of the frameworks can have an impact on the results. We attempted to configure each framework for best efficiency (mainly at the level of memory management, parallelism and processor setup), but an exhaustive search of all best

configurations would be unfeasible. Our scenario was more exploratory with respect to power consumption anomaly detection. A full experiment would need to take into account changes to parameters and the impact on the performance. For example, the number of deployed nodes alone can have a different impact on each of the considered frameworks. In our case, we kept a rather simple nodes topology, but for more complex topologies the results can be different (as previous research has shown, e.g., [37]). Another internal threat to validity is given by the implementation differences of the algorithm on each platform. Each framework provides different abstractions to develop applications and it is not possible to implement the algorithm equally, although we believe this threat is limited due to the simple anomaly detection algorithm we applied for benchmarking. Another threat is related to construct validity, the scenario was meant to compare the frameworks in a common data processing context and not to be a full experiment in which many factors are varied, like the number of nodes to which each framework is distributed. Another threat is related to generalization, the results apply to the specific scenario discussed to showcase the platform, other scenarios might have other needs and lead to different results.

**Frameworks versions.** In running the power consumption anomaly detection scenario, the following frameworks versions have been used:

Apache Zookeeper 3.4.12, Apache Kafka 2.0.0,  
Apache Hadoop 2.8.5, Apache Flink 1.7.1,  
Apache Spark 2.4.1, Apache Storm 1.2.2,  
Scala 2.12, Java JDK 1.8.0201

## VI. CONCLUSION

Big data processing in the Smart Grids context has many applications that require real-time operations and stream processing. In this paper, we presented a big data platform for anomaly detection from power consumption data. The platform is based on an ingestion layer with data densification, Apache Flink as part of the speed layer and HDFS/KairosDB as the data store. We mapped the main components to the reference architecture proposed by Pääkkönen and Pakkala [12], and provided the results of a scenario based on power consumption anomaly detection to assess the applicability of different frameworks: batch-based (Spark), stream-based (Storm), or hybrid (Flink). Overall, we adopted Flink in the platform's speed layer, as it provided the best performance for stream processing and met the requirements for power consumption datasets anomaly detection in our scenario.

Currently, we are planning to deploy the platform to analyze large-scale power consumption datasets in our running projects, by comparing several anomaly detection algorithms to help in better identifying clusters of customers based on smart metering data traces.

## ACKNOWLEDGMENT

The work was supported from European Regional Development Fund Project *CERIT Scientific Cloud* (No.

CZ.02.1.01/0.0/0.0/16\_013/0001802). Access to the CERIT-SC computing and storage facilities provided by the CERIT-SC Center, provided under the programme "Projects of Large Research, Development, and Innovations Infrastructures" (CERIT Scientific Cloud LM2015085), is greatly appreciated.

## REFERENCES

- [1] Y. Demchenko, P. Grosso, C. De Laat, and P. Membrey, "Addressing big data issues in scientific data infrastructure," in *2013 International Conference on Collaboration Technologies and Systems (CTS)*. IEEE, 2013. doi: 10.1109/CTS.2013.6567203 pp. 48–55.
- [2] A. Radenski, T. Gurov, K. Kaloyanova, N. Kirov, M. Nisheva, P. Stanchev, and E. Stoimenova, "Big data techniques, systems, applications, and platforms: Case studies from academia," in *2016 Federated Conference on Computer Science and Information Systems (FedCSIS)*. IEEE, 2016. doi: 10.15439/2016F91 pp. 883–888.
- [3] Y. Zhang, T. Huang, and E. F. Bompard, "Big data analytics in smart grids: A review," *Energy Informatics*, vol. 1, no. 1, p. 8, 2018. doi: 10.1186/s42162-018-0007-5
- [4] K. Zhou, C. Fu, and S. Yang, "Big data driven smart energy management: From big data to big insights," *Renewable and Sustainable Energy Reviews*, vol. 56, pp. 215–225, 2016. doi: 10.1016/j.rser.2015.11.050
- [5] B. Rossi and S. Chren, "Smart grids data analysis: A systematic mapping study," *arXiv preprint arXiv:1808.00156*, 2018. [Online]. Available: <https://arxiv.org/abs/1808.00156>
- [6] M. Ge, H. Bangui, and B. Buhnova, "Big data for internet of things: a survey," *Future Generation Computer Systems*, vol. 87, pp. 601–614, 2018. doi: 10.1016/j.future.2018.04.053
- [7] X. Liu and P. S. Nielsen, "Regression-based online anomaly detection for smart grid data," *arXiv preprint arXiv:1606.05781*, 2016.
- [8] B. Rossi, S. Chren, B. Buhnova, and T. Pitner, "Anomaly detection in smart grid data: An experience report," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2016. doi: 10.1109/SMC.2016.7844583 pp. 2313–2318.
- [9] Z.-H. Yu and W.-L. Chin, "Blind False Data Injection Attack Using PCA Approximation Method in Smart Grid," *IEEE Transactions on Smart Grid*, vol. 6, no. 3, pp. 1219–1226, 2015. doi: 10.1109/TSG.2014.2382714
- [10] J. Lee, Y.-c. Kim, and G.-L. Park, "An Analysis of Smart Meter Readings Using Artificial Neural Networks," in *Convergence and Hybrid Information Technology*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, vol. 7425, pp. 182–188.
- [11] F. McLoughlin, A. Duffy, and M. Conlon, "A clustering approach to domestic electricity load profile characterisation using smart metering data," *Applied energy*, vol. 141, pp. 190–199, 2015. doi: 10.1016/j.apenergy.2014.12.039
- [12] P. Pääkkönen and D. Pakkala, "Reference architecture and classification of technologies, products and services for big data systems," *Big data research*, vol. 2, no. 4, pp. 166–186, 2015. doi: 10.1016/j.bdr.2015.01.001
- [13] N. Marz and J. Warren, *Big Data: Principles and best practices of scalable real-time data systems*. New York; Manning Publications Co., 2015.
- [14] J. Lin, "The lambda and the kappa," *IEEE Internet Computing*, vol. 21, no. 5, pp. 60–66, 2017. doi: 10.1109/MIC.2017.3481351
- [15] S. Shahrivari, "Beyond batch processing: towards real-time and streaming big data," *Computers*, vol. 3, no. 4, pp. 117–129, 2014. doi: 10.3390/computers3040117
- [16] J. Kreps, "Questioning the lambda architecture," *Online article, July*, 2014. [Online]. Available: <https://www.oreilly.com/ideas/questioning-the-lambda-architecture>
- [17] M. Kiran, P. Murphy, I. Monga, J. Dugan, and S. S. Baveja, "Lambda architecture for cost-effective batch and speed big data processing," in *2015 IEEE International Conference on Big Data (Big Data)*. IEEE, 2015. doi: 10.1109/BigData.2015.7364082 pp. 2785–2792.
- [18] J. van Rooij, V. Gulisano, and M. Papatriantafyllou, "Locovolt: Distributed detection of broken meters in smart grids through stream processing," in *Proceedings of the 12th ACM International Conference on Distributed and Event-based Systems*. ACM, 2018. doi: 10.1145/3210284.3210298 pp. 171–182.

- [19] H. Sequeira, P. Carreira, T. Goldschmidt, and P. Vorst, "Energy cloud: Real-time cloud-native energy management system to monitor and analyze energy consumption in multiple industrial sites," in *2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing*. IEEE, 2014. doi: 10.1109/UCC.2014.79 pp. 529–534.
- [20] M. Mayilvaganan and M. Sabitha, "A cloud-based architecture for big-data analytics in smart grid: A proposal," in *2013 IEEE International Conference on Computational Intelligence and Computing Research*, Dec 2013. doi: 10.1109/ICCIC.2013.6724168 pp. 1–4.
- [21] A. A. Munshi and Y. A. I. Mohamed, "Data lake lambda architecture for smart grids big data analytics," *IEEE Access*, vol. 6, pp. 40463–40471, 2018. doi: 10.1109/ACCESS.2018.2858256
- [22] X. Liu and P. Nielsen, "Streamlining smart meter data analytics," in *Proceedings of the 10th Conference on Sustainable Development of Energy, Water and Environment Systems*. International Centre for Sustainable Development of Energy, Water and Environment Systems, 2015.
- [23] N. Balac, T. Sipes, N. Wolter, K. Nunes, B. Sinkovits, and H. Karimabadi, "Large scale predictive analytics for real-time energy management," in *2013 IEEE International Conference on Big Data*, Oct 2013. doi: 10.1109/BigData.2013.6691635 pp. 657–664.
- [24] A. R. Al-Ali, I. A. Zualkernan, M. Rashid, R. Gupta, and M. Alikarar, "A smart home energy management system using iot and big data analytics approach," *IEEE Transactions on Consumer Electronics*, vol. 63, no. 4, pp. 426–434, November 2017. doi: 10.1109/TCE.2017.015014
- [25] H. Daki, A. El Hannani, A. Aqqal, A. Haidine, and A. Dahbi, "Big data management in smart grid: concepts, requirements and implementation," *Journal of Big Data*, vol. 4, 12 2017. doi: 10.1186/s40537-017-0070-y
- [26] C. Yang, W. Chen, K. Huang, J. Liu, W. Hsu, and C. Hsu, "Implementation of smart power management and service system on cloud computing," in *2012 9th International Conference on Ubiquitous Intelligence and Computing and 9th International Conference on Autonomic and Trusted Computing*, Sep. 2012. doi: 10.1109/UIC-ATC.2012.160 pp. 924–929.
- [27] M. Ali, Z. Y. Dong, X. Li, and P. Zhang, "Rsa-grid: a grid computing based framework for power system reliability and security analysis," in *2006 IEEE Power Engineering Society General Meeting*, June 2006. doi: 10.1109/PES.2006.1709374. ISSN 1932-5517
- [28] T. Rajeev and S. Ashok, "A cloud computing approach for power management of microgrids," in *ISGT2011-India*, Dec 2011. doi: 10.1109/ISET-India.2011.6145354 pp. 49–52.
- [29] S. Chren, B. Rossi, and T. Pitner, "Smart grids deployments within eu projects: The role of smart meters," in *2016 Smart cities symposium Prague (SCSP)*. IEEE, 2016. doi: 10.1109/SCSP.2016.7501033 pp. 1–5.
- [30] M. Schvarcbacher, K. Hrabovská, B. Rossi, and T. Pitner, "Smart grid testing management platform (sgtmp)," *Applied Sciences*, vol. 8, no. 11, p. 2278, 2018. doi: 10.3390/app8112278
- [31] K. Hrabovská, N. Šimková, B. Rossi, and T. Pitner, "Smart grids and software testing process models," in *2019 Smart cities symposium Prague (SCSP)*. IEEE, 2019, pp. 1–5.
- [32] J. Peppanen, X. Zhang, S. Grijalva, and M. J. Reno, "Handling bad or missing smart meter data through advanced data imputation," in *IEEE Innovative Smart Grid Technologies Conference*. IEEE, 2016. doi: 10.1109/ISGT.2016.7781213 pp. 1–5.
- [33] X. Liu, N. Iftikhar, H. Huo, R. Li, and P. S. Nielsen, "Two approaches for synthesizing scalable residential energy consumption data," *Future Generation Computer Systems*, vol. 95, pp. 586–600, 2019. doi: 10.1016/j.future.2019.01.045
- [34] M. Zaharia, R. S. Xin, P. Wendell, T. Das, M. Armbrust, A. Dave, X. Meng, J. Rosen, S. Venkataraman, M. J. Franklin *et al.*, "Apache spark: a unified engine for big data processing," *Communications of the ACM*, vol. 59, no. 11, pp. 56–65, 2016. doi: 10.1145/2934664
- [35] P. Carbone, A. Katsifodimos, S. Ewen, V. Markl, S. Haridi, and K. Tzoumas, "Apache flink: Stream and batch processing in a single engine," *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, vol. 36, no. 4, 2015.
- [36] S. T. Allen, M. Jankowski, and P. Pathirana, *Storm Applied: Strategies for real-time event processing*. Manning Publications Co., 2015.
- [37] J. Karimov, T. Rabl, A. Katsifodimos, R. Samarev, H. Heiskanen, and V. Markl, "Benchmarking distributed stream processing engines," *arXiv preprint arXiv:1802.08496*, 2018.
- [38] Y. Wang, "Stream processing systems benchmark: Streambench," G2 Pro gradu, diplomityö, Aalto University, Finland, 2016-06-13. [Online]. Available: <http://urn.fi/URN:NBN:fi:aalto-201606172599>
- [39] M. A. Lopez, A. G. P. Lobato, and O. C. M. Duarte, "A performance comparison of open-source stream processing platforms," in *2016 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2016. doi: 10.1109/GLOCOM.2016.7841533 pp. 1–6.
- [40] S. Chintapalli, D. Dagit, B. Evans, R. Farivar, T. Graves, M. Holderbaugh, Z. Liu, K. Nusbaum, K. Patil, B. J. Peng *et al.*, "Benchmarking streaming computation engines: Storm, flink and spark streaming," in *2016 IEEE international parallel and distributed processing symposium workshops (IPDPSW)*. IEEE, 2016. doi: 10.1109/IPDPSW.2016.138 pp. 1789–1792.
- [41] Smart\* data set for sustainability. [Online]. Available: <http://traces.cs.umass.edu/index.php/Smart/Smart>

# Search for the Memory Duplicities in the Java Applications Using Shallow and Deep Object Comparison

Richard Lipka

NTIS - New Technologies for the Information Society  
Faculty of Applied Sciences  
University of West Bohemia  
Univerzitni 8, Plzen, 323 00, Czech Republic  
Email: lipka@kiv.zcu.cz

Tomas Potuzak

Department of Computer Science and Engineering  
Faculty of Applied Sciences  
University of West Bohemia  
Univerzitni 8, Plzen, 323 00, Czech Republic  
Email: tpotuzak@kiv.zcu.cz

**Abstract**—In high-level object languages, such as Java, a problem of unnecessary duplicates of instances can easily appear. Although there can be a valid reason for maintaining several clones of the same data in the memory, often it indicates that the application can be refactored into a more efficient one. Unnecessary instances consume memory, but in case of Java applications can also have a significant impact on the application performance, as they might prolong the time needed for the garbage collection. In this paper, we are presenting a method and a tool that allows detecting duplicity in the heap dump of a Java application, based on the shallow and deep object comparison. The tool allows to identify the problematic instances in the memory and thus helps programmers to create a better application. On several case studies, we also demonstrate that the duplicates appear not only in the student projects and similar programs that often suffer from poor maintenance but also in commonly available Java tools and frameworks.

## I. INTRODUCTION

JAVA language was designed to provide fully automated memory management and to shield programmers from errors caused by the memory leaks. The developers are often encouraged to design the data models based on the real world structures and not to think too much about the internal representation of the data they are using. This should lead to greater efficiency of programming, but at the same time hardware resources are often used inefficiently. Consequently, instead of memory leaks typical in C-language programs, programmers are creating different constructions that clog up the operational memory and can lead to unnecessary slowdowns of the application (or even to the termination of the application due to insufficient memory) due to excessive garbage collection.

In general, this problem is known as the memory bloat, and there are many different aspects of it [1]. We had previous experience with fixing an application that was suffering heavily from wasting memory [2], so our main goal was to create a tool that would allow us to easily identify the problematic objects retained in the Java heap. One of the issues we have encountered is (often multiple) duplication of the identical instances in the memory. Especially for less

experienced programmers, it might be difficult to identify the problem. We hope that our tool might help them to find the unnecessary objects in the memory of their programs. In the same time, we wanted to investigate how often the similar problem arises in other applications generally available in the Java community.

### A. Memory Bloat

There is no generally accepted definition of the memory bloat, but many examples are known both from the literature and from the real applications. In [1] Mitchell describes 15 anecdotal examples of the memory issues that might arise, classified into four main groups, and shows how Java Virtual Machine deals with them and how programmers might or might not make its work more difficult. Anecdote 12 mentions data duplication created during the communication between Java application and the outside environment. Mitchell also describes how different types of objects can have a significant impact not only on the memory consumption but also on the application speed, as the Java Virtual Machine has to perform garbage collections. Depending on the number of retained objects, it can significantly slow down the application.

The problem of the object duplication is partially solved also in Java Virtual Machine itself, currently only with the `String` class. Since Java version 8.20 [3] Java contains an implementation of the string deduplication – as long as `Strings` are managed by the virtual machine, they are created in the separate part of the memory. When a new `String` shall be created, it is first checked whether there already is an instance with the same content. If so, only a reference on the existing instance is provided. As `Strings` are immutable in Java, this is a safe way of dealing with them - all operations manipulating with `Strings` are in fact creating new instances, so when a programmer wants to change the content of the instance, a new, different instance with new data is provided.

This behaviour is possible mainly due to the simple nature of the `String` object – only an array of characters needs to be checked. However, not only `Strings` instances are

duplicated in Java programs. Our intention is not to provide a more general, runtime method for the deduplication, as the deep object comparison can be quite time-consuming. We only intend to provide a tool that will allow to analyze the program memory and to discover possible duplicates. We also do not claim that the duplicate objects can be automatically merged just because they contain the same data - this decision has to be made by a programmer with a deep insight into the application. However, if there are multiple identical instances of one class, it can be a strong indicator that the application can be refactored into a more efficient one.

### B. Motivation

Our motivation comes from two sources. First, we are often dealing with software created by students, which usually contains many types of problems and we were not able to find a tool that would be able to show how many duplicates are present in the memory of student programs. Standard profilers such as *VisualVM* can be helpful, but not suitable for this type of analysis. Furthermore, we wanted to see if this problem is present not only in the work of inexperienced programmers but also in software that is more widespread and freely available.

The remainder of the paper is organized as follows: Section II deals with the related work, focused mainly on the problem of the equality and comparison of the objects. Section III explains how the equality of the objects is implemented in our tool. In section IV we are showing the algorithm used for the duplicity analysis in our implementation. The method of validation of our implementation, as well as several results of duplication analysis of several Java application, is presented in section V. The last section concludes the paper and discusses possible future work.

## II. RELATED WORK

There are two main areas relate to our work. The first one is the problem of the memory efficiency of Java applications. This problem is discussed quite extensively and many different bad practices or problematic patterns were described in last years. Especially with the increasing interest in the embedded systems, the need for the memory efficient software grows [4].

The most common problem is the detection of the memory leaks, described for example in [5] or in [6]. As the Java is a language with garbage collection, the classical memory leaks with inaccessible memory are rare in it, and the works we are mentioning are focusing more on the detection of the objects with large overhead [5]. The typical example might be collections that contain mostly `null` elements. Another approach is the design of the more efficient ways of the garbage collection [6]. As one of the main sources of the performance issues and memory wasting is automated ORM (Object-Relational Mapping) when used incorrectly, it is also possible to find approaches focusing on the analysis of the ORM performance antipatterns and fixing them [7].

Other approaches are trying to find a way how to evaluate overall memory health of the Java programs. One of the most

useful works on this topic is [8], where the problematic structures are described in high detail and even a metric based on the ratio of the useful data and structure overhead is proposed to evaluate memory health. These ideas are further expanded in [1] and [9], where the memory impact of the complexity of the domain model is discussed, as each reference occupy some space. Several different antipatterns are presented, along with the proposed solutions. Unfortunately, the patterns described in this work are not discovered automatically, the authors rely mostly on the manual analysis of the data structures. Another approach to the memory optimization is based on the efficient memory partitioning [10].

One important inspiration for our work is [11], where both static and dynamic analysis is proposed as a tool for evaluating the usage of Java collections (or other structures that allows manipulation with a large number of elements). Another approach for the dynamic analysis of the collections efficiency is presented in the [12] – the technique proposed here aims not only to use collections more efficiently, but also to choose the most suitable collection for the application, based on the runtime analysis.

The second is more focused on the instances comparison or even automatic detection of the possibility of replacing one instance with another. In [13], a post-mortem analysis of the Pharo programs is proposed in order to determine if there are some redundancies in the suitable classes with inclination to the redundancy (such as `Point` or `String` classes), aiming to replace the redundant ones with one instance. This paper also contains an extensive description of different ways how to define an object equivalence. The possibility of the replacement of one instance with another is also examined in [14]. In this case, not only a comparison of the objects is performed, but authors also propose to instrument the original program in order to observe the usage of the candidates for the merging and automatically determine if such merging is possible without influencing the program behaviour.

As the search for duplicates in the whole namespace is time demanding task with a high complexity, some publications are focusing only on the classes which are known to contain duplicates very often. In [15], the methods used for the `String` deduplication is described in high detail. Similarly, the description of the approaches to get rid of the `String` duplicates is described in [3]. As this is typically performed at runtime, there is a great need to make these algorithms as efficient as possible. In [16], different methods for the decision whether the deduplication should be performed or not and their impact on the program performance is demonstrated.

## III. OBJECT EQUALITY

As was mentioned in previous section, there are multiple ways of how the object equality can be defined. As our application is working with the heap dump, we did not focus on the fast pre-analysis using some form of the hash code of the objects, such as in [17], but immediately on the analysis of the attributes of the objects, similar as described in [18]. In contrast with [18], we are not working directly with objects

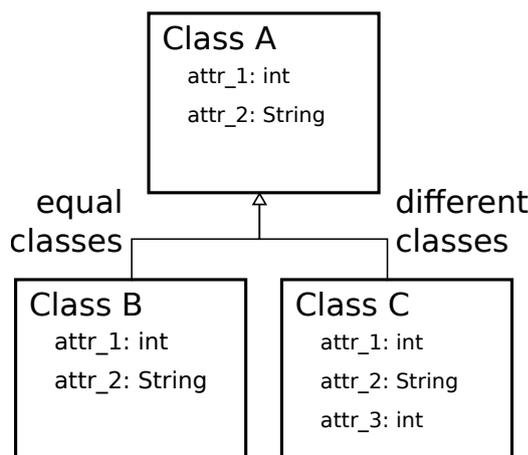


Fig. 1. Class equality in inheritance chain

in memory, but with their serialized form in the heap dump. So for the comparison, we cannot use in any way methods that are implemented in the objects, such as `equals()` or `compareTo()`.

#### A. Class Equality

The first aspect to consider is the class of the compared objects. If we are comparing instances with the same class, it makes sense to analyze the data field by field. If the values of all fields are identical, we can consider the instances identical as well. However, it is also possible to consider inheritance. The inheritance in Java cannot remove any field from the successor, only to add additional ones or to add or override its methods. In case that the predecessor contains the same fields as the successor, we can consider them identical from the data point of view as well. Of course, there is a question if such objects can be merged into one, but that is something that can not be decided automatically during the heap analysis. Thus, we consider instances identical in case all their fields are identical even when they belong to different classes, as long as they are part of the same inheritance chain and no additional fields are added in the successors (as you can see in Fig. 1). When such field is added, even if its value is `null`, it is considered different due to the different class definition and the matching of fields is not performed at all. You can notice that including inheritance chains in the comparison, in fact, broadens the definition of what are identical objects and can lead to a higher number of identified duplicates.

#### B. Fields Equality

When fields are compared, it is simple to deal with the primitive types, but more complicated to deal with the references (see Section III. C). Although `Strings` are references, Java deals with them in a special way. This enables us to treat them in a special way as well. As they are stored in a special part of memory and due to the string deduplication, we often do not need to analyze the actual content of the `String`, only to see if the reference leads to the same instance. When

references are different, the actual data of the `String` have to be compared. This can happen, as the string deduplication does not work for all `Strings` in Java application and the instance of the `String` can thus appear both in the regular heap space and in the area reserved for the `Strings`.

#### C. References Equality

Another aspect is dealing with the reference fields. The previously described method is suitable for the objects that are composed only of the primitive data types and `Strings`. However, in Java, most objects contain also references on other objects. In such a case, there are two different points of view. The shallow approach would consider two references identical only when they are pointing to the same instance. This can be checked very easily, as in the heap dump, the references are represented only as a `long` number, so we only need to compare those.

In order to obtain a broader set of results, we have also implemented a deep comparison approach. It means that, in addition to the identical reference numbers, two references will be considered identical when they point to two different instances that are internally identical. Again, this approach leads to a higher number of identified duplicates, but also significantly increases the complexity of the comparison, as it has to be used in a recursive way – the referred instances can point to other instances and so on. This also means that a stopping condition has to be defined, in order to deal with cycles and to improve the performance of the comparison.

The references can create an arbitrary graph, but it can be always reduced to a finite tree structure when the analysis is stopped after each node of the graph is visited once. Another option is to define a depth, in which the analysis should end. If the instances fields contents are identical till the required depth is reached, the instances themselves are considered identical as well.

The last aspect we need to describe is the dealing with arrays and collections. Java offers `List` and `Set` interfaces and several implementations that can be used to store a large number of data. In case when the deep object comparison is used, no special approach is needed and the structures are identical only when they contain identical instances in the same order. It would be possible to broaden equality definition even more and ignore the order of the instances within the array, but that would require even more complex calculation and specific implementation for each Java collection.

## IV. DUPLICATION FINDER

Our implementation of the duplication finder is created in the Java language. As we need to deal with a heap dumps in binary HPROF format, which is created by means of JVM, we were looking for a tool that would allow us to process the data easily and we used a Hprof Heap Dump parser library [19]. This library allows us to load the data from heap dump and reconstruct the content of each instance. It also provides access to the class descriptions so the data can be correctly interpreted.

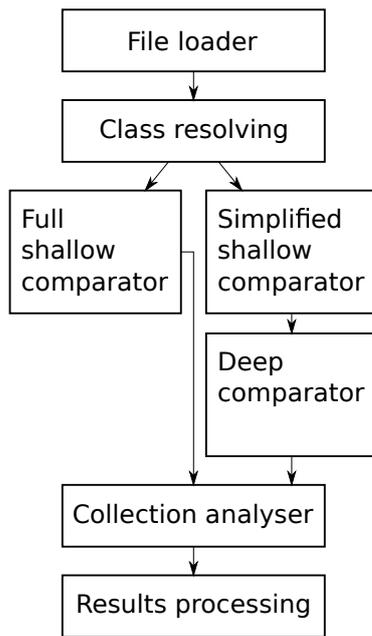


Fig. 2. Tool pipeline

### A. Tool Architecture

In order to allow easy modification of the analysis process (for example to be able to switch between shallow and deep comparison or to add additional modules searching for other memory anti-patterns), the tool is designed as a pipeline. Our tool sequentially reads the heap dump file and produce a basic representation of the loaded data. The overall behaviour is represented in Fig. 2.

These data are joined with the corresponding class descriptions and then further processed according to the class description (so the loaded byte streams are converted for example to long numbers or to `Strings` for easier processing). When the actual class of each instance is analyzed, it is also possible to decide if the analysis should stop or continue depending on the class or package name – this allows us to limit the duplicity search only on the certain classes in the memory dump and thus save some time during the analysis.

### B. Problem Identification

The prepared data are then passed to a module that is responsible for the duplicity analysis. The matching algorithm iterates over all loaded instances and stores them according to their properties in the two-level structure. At first, instances are divided according to their classes (as was described in the previous section – so the different classes can be considered equivalent if they are part of the same inheritance chain and they share the same set of fields). Then, within each class, the equivalent instances are grouped according to the values of their fields (see Fig. 3). Depending on the settings, the shallow or the deep comparison of the objects is performed in this phase.

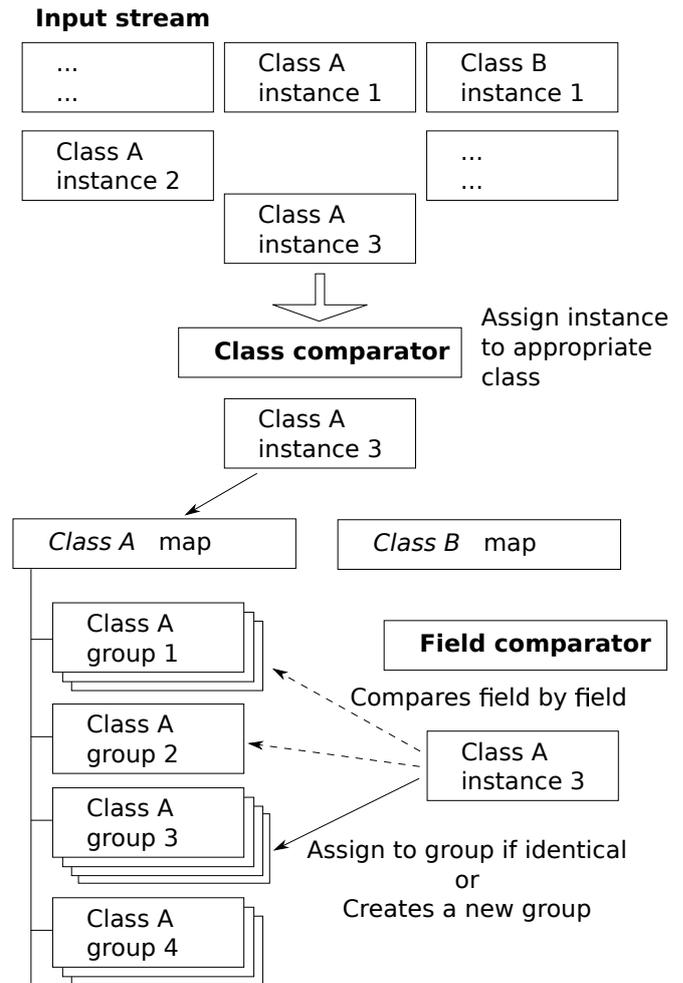


Fig. 3. Data structures for the equality classification

The shallow comparison is quite straightforward – the instances retrieved from the heap dump are compared field by field, including the references (see Fig. 4). If the instances are identical on this level, they are considered equal and became part of the group within the class. Each additional instance is compared against all existing groups and either is added to one of the groups or a new group is created if the new instance is unique.

Deep comparison is more complicated and more time demanding. The algorithm is similar to the deep object comparison algorithm we have described in [20], but for this purpose modified and made faster. There are two main differences – the first one is that we are now working with the heap dump data and not with the instances that are in memory of the executed application. The second is that we do not need to explore the structures of instances completely - the first occurrence of a difference is sufficient to claim that the compared objects are not identical. It would be possible to use the original version of the algorithm as well, but it would make the whole comparison process even more time demanding.

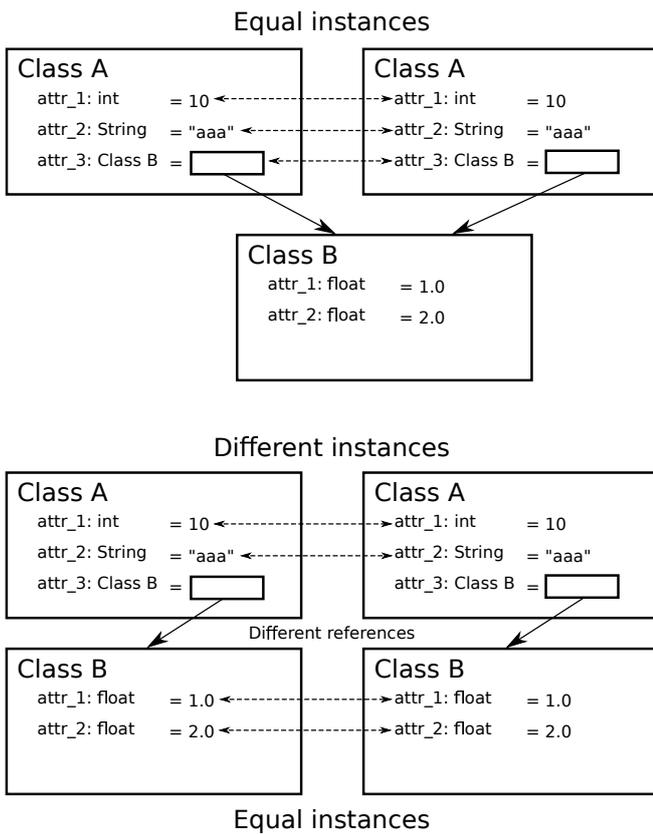


Fig. 4. Instances equality with Shallow comparison

In order to perform the deep comparison, a graph representation of both compared objects has to be created. In order to do so, the whole memory dump have to be processed (we have to be able to resolve references to create an object graph), so it is performed after the shallow comparison is finished. But for the purpose of the deep comparison, a modified version of shallow comparison is done. In this case, only primitive data types are compared to determine if the instances are identical and the references are ignored. This, of course, means that the instances that differ only in references will be considered identical during this modified shallow comparison. The reason for this is to allow faster evaluation of the deep comparison. These "duplicates" are not reported in the result of the algorithm, but only used in order to evaluate the equality of the referred instances during the deep comparison faster.

Each node of the graph corresponds to an instance and each edge corresponds to the reference. As we expect that, in most cases, the instances will not be identical, the graph is constructed on demand, as a modification of Breadth First Search (BFS) algorithm. When all fields of compared instances are identical, the references are resolved one by one (see Fig 5). As the deep comparison is actually performed after the shallow one, in the first step algorithm checks if the referred objects were identical during the shallow comparison. If not, the comparison can be terminated, as we are not looking for all

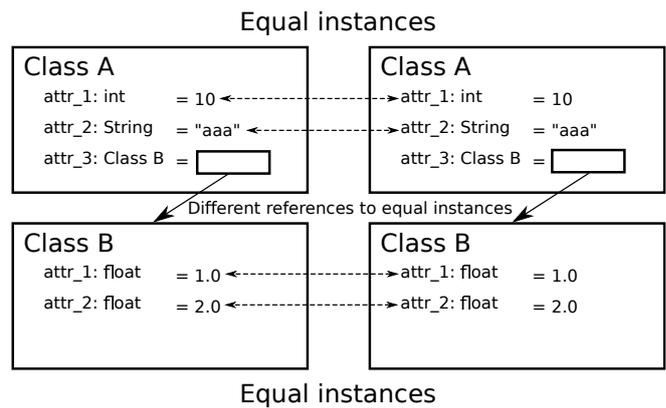


Fig. 5. Instances equality with Deep comparison

differences between instances, only one difference is sufficient to consider them unique. If the referred instances belong to the same shallow equivalent class, other references of the original object are analyzed. Only when all referred instances are considered identical, the algorithm starts constructing the next level of the graph.

Finally, the module for the analysis of the lists and arrays is used. For this analysis, we considered using the idea of collection health described for example in [8]. Collections are considered to be healthy (from the memory utilization point of view) only if they carry a sufficient amount of data. Programmers in Java often create unnecessary large pre-allocated structures, which are never sufficiently utilized, so the collection contains a significant number of null references. Such references, despite not carrying any data, still occupy memory space. For example, default instance of the ArrayList is created with 10 empty slots and often only small fraction is used (more specifically, the array is allocated when the first element is inserted to it, when the collection is created empty, the array is not created immediately).

This analysis is currently limited only on the offsprings of the AbstractList class, which are based on the array. As each collection requires a different approach and implementation, we focused only on the array types and not on the linked structures or maps. The analytical module is looking for two parameters. First, it calculates the ratio of the space occupied by the collection to its size. All collections that have this ratio below 0.5 are marked as underutilized. Second, the content is analyzed for duplicity and, if the collection is filled with identical elements, it is marked as problematic as well.

### C. Result Reporting

The last part of the processing is the results reporting. As the tool is so far only operated from the command line, the results are presented only in the text form in the standard output stream. The result report contains names of all classes that have at least one set of duplicate instances (including information about the package and the inheritance chain), the serialized form of fields of duplicate instances and also the list of collections that contain one of the marked problems.

#### D. Complexity

The complexity of comparing each instance with each other to determine if they are identical is implemented with quadratic complexity with respect to the number of elements to compare. It can be expressed as  $O(n^2)$ , where  $n$  is a number of elements to compare. However we do not really need to compare each element with each other across the whole heap dump - we need to compare elements only within the same class. In such case, complexity is still quadratic, but in a form  $O(k \cdot n_k^2)$  where  $k$  is the number of equivalent classes and  $n_k$  is a number of the instances within each class. This makes the comparison more feasible, the number of instances within a class is more manageable value. We have to note here that the time even for the shallow comparison itself can differ significantly according to the number of the fields in compared classes. In case of the deep comparison, it depends heavily on the complexity of the compared structures. This has a significant impact on algorithm performance as well. Theoretically, it would be possible to achieve linear complexity by calculating hash for each instance and compare only the hash codes, but the calculation of hash for the deep comparison would still require a complete reconstruction of the data structures, even in cases when they differ in first few attributes (and the comparison will quickly find differences), so we decided against this approach.

### V. VALIDATION AND RESULTS

In order to validate the functionality of our tool, we have at first created a simple test application with a known number of duplicates and half empty collections. The purpose was just to figure out whether the tool will be able to find all injected problems. Then we have continued with tests performed on several real-life applications, to investigate whether the duplicates occur in the real world software.

#### A. Testing Application

The testing application is able to generate an arbitrary number of instances of simple objects, containing numerical and string attributes, as well as simple reference structures. It uses only two simple classes, `Child` and `Parent` that can refer to each other. The testing application was run several times with the different number of the created instances, in order to verify the time complexity of the algorithm. In each test, there were only 5 deep duplicates.

The heap dump was obtained using `jmap` [21] tool, using command

```
jmap -dump:live , file=<file -path> <pid>
```

in order to ensure that only "living" objects (objects that would survive next garbage collection) are obtained and listed in the heap dump. The time measurements were done on the PC with Windows 10, SSD drive, Intel Core i7-4930K CPU, 3.40 GHz and 32 GB RAM. No parallelization was used at the moment. The analysis was limited only to the package with our data classes, other instances were ignored. The times were obtained as an average from 5 executions.

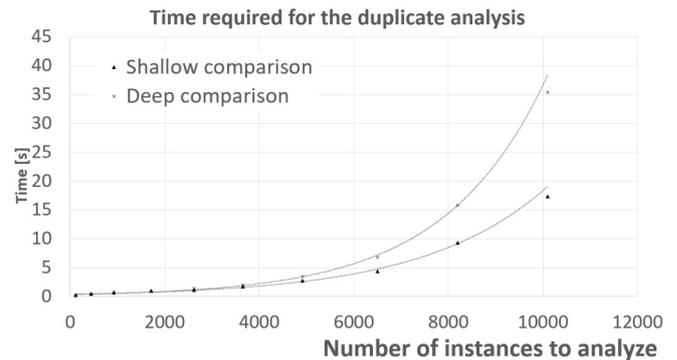


Fig. 6. Times required for analysis by deep comparison and shallow comparison algorithms

In Table I, the results for the testing application are presented. From these results, the deep and shallow comparisons can be compared both in their abilities and consumed computation time. In this simple case, our deep comparison algorithm was able to find all duplicates, while the shallow comparison version was not able to identify duplicates that were using different referred instances with identical data. On the other hand, the time required for the deep comparison grows significantly faster even when there are no significantly more deep duplicates (see Fig. 6). Running the deep analysis thus makes sense only if there is a strong suspicion that there might be such kind of problem. As expected, time requirements appear to grow quadratic in relation to the number of instances that are there for comparison, although the measurements, especially for the small number of instances might be influenced by the overhead of the processing.

#### B. Real World Programs

We have also attempted to run the algorithm on several real-world programs, to investigate whether if the problem with memory duplicates will be present in them. We have chosen four different programs – *Spring Boot framework* with a basic hello-world application, *Eclipse* with several projects open, *IntelliJ Idea* and *TomEE* with running map visualization server. Due to the time complexity of the deep analysis, we used only a shallow comparison in all four examples.

1) *Spring Boot*: The *Spring Boot framework* [22] in version 2.1.4 was used with only the very basic "Hello World" application. The dump of the whole framework has approximately 27 MB. For the purposes of the experiment, we have decided to limit our testing only to the classes from the `org` namespace.

In Table II, the results of the `org` namespace and its parts are summarized. One important find is that some duplicates were in fact found, but there is not a great number of them. Mostly, the duplicities were only pairs, but for example, the `Signature` class had 38 identical instances and class `DefaultFlowMessageFactory` 34 instances. Both classes only contain short `Strings` with basic framework settings, but their presence in the analyzed application shows that our tool is able to work with the real software.

TABLE I  
RESULTS FOR THE SIMPLE TESTING APPLICATION

Instance count	Dump size [MB]	Injected duplicates	Found duplicates (shallow)	Found duplicates (deep)	Duration shallow [ms]	Duration deep [ms]
120	2.1	20	15	20	221	227
440	2.6	30	25	30	493	401
920	3.0	40	35	40	725	705
1710	3.6	50	45	50	987	912
2620	4.0	60	55	60	1121	1409
3650	4.4	70	65	70	1753	2021
4910	5.4	80	75	80	2769	3517
6500	5.9	90	85	90	4379	6781
8200	6.4	100	95	100	9334	15789
10100	6.9	110	105	110	17351	35419

TABLE II  
ANALYSIS OF CLASSES IN SPRING BOOT FRAMEWORK

Package name	Classes	Instances	Found duplicates	Duration [ms]
org	2416	9093	347	14759
org.springframework	1555	6053	329	8214
org.springframework.boot	380	1506	27	4229
org.springframework.core	196	1585	5	4425
org.springframework.web	296	239	37	4108
org.springframework.boot.web	75	27	1	4002

Furthermore, as the Spring Boot framework is often used and well maintained, we did not expect to find many problems in it.

The second thing to notice is the ratio of classes and instances within one class. In the whole `org` package the ratio is approximately 1 : 4. For many classes, there are only a few instances. During drill down to the sub-packages, the ratio changes. For example in the package `org.springframework.boot.web` the ratio is even reversed – more classes were loaded from the namespace than was actually used to create instances.

2) *Eclipse*: We have analyzed *Eclipse* [23] in version 4.10.0 (build 20181214-0600). The IDE was only started and in the moment of heap dump collecting was not performing any particular task. The size of the Eclipse heap dump was approximately 92 MB. Measurements were performed for the packages `org`, `com`, `java`, `sun`, and `ch`. The results are summarized in Table IV.

This is the largest heap we have processed and, again like in the heap of the *Spring framework*, no large problem is present. However, the tool demonstrates that it is capable handling not only trivial examples but also larger datasets. The most duplicated was the class `org.eclipse.swt.widgets.TypedListener` with 444 identical instances based on the shallow comparison. Many of the discovered classes contained large fragments of the XML configuration of the tool (like `org.eclipse.sisu.plexus.ConfigurationImpl`

with 16 identical instances containing 750 characters each). The results also show the rapid growth of the required computational time for larger datasets, with the analysis of `java` package taking more than 6 hours. In this package, 75 ms on average was required for analysis of each instance. In comparison in the smallest package `ch`, only 13 ms on average were required.

3) *IntelliJ Idea*: Along with Eclipse, we also tried to perform analysis of IntelliJ Idea in version 2018.3. The dump of this IDE was smaller than in the case of the Eclipse, approximately 74 MB. Only packages `org`, `com`, and `sun` were available within. In a similar way as in the previous case, the IDE was not performing any particular task, it was just started. The results are summarized in the Table IV.

In this case, despite the lower number of the classes with duplicates, a large number of identical instances was found. The class `org.jdom.Text` contained several instances with many clones, the largest group had 11577 identical instances. All these clones contained only several unprintable characters (typically end of the line and some other character) and obviously were part of the loaded DOM of some data the IDE was requiring after starting. In this case, the tool demonstrates that it is capable of discovering large clusters of the identical instances. However further analysis of the source texts of the *IntelliJ Idea* would be required to determine if there is a way to mitigate this type of the duplicity. Other duplicity classes (with only several clones) contained for example the text of the library licenses.

4) *TomEE with Visualisation Server*: The last example we tried to analyze was Apache *TomEE* [22] server in version 7.1.5, with the running application dealing with the visualization of the power grid. *TomEE* is a version of the Tomcat server, with additional modules useful for building enterprise applications. In this case, we have decided to focus not on the classes from the technology itself, but on the domain objects from the visualization server. As previous examples showed, the frameworks that are intensively used will probably contain fewer issues than the applications that should be executed within them.

The size of the heap dump of the server was approximately 370 MB, significantly larger than the previous files. When the dump was collected, the server was working with 4 users at the moment, so 4 sessions with data models were loaded. We were focused on the proprietary package `cz.zcu.laps.pnp.domain`, which contains domain data of the application. The data were organized in the form of a graph, composed of the nodes representing elements of the power distribution network and power lines between them. Each user is able to work only with one model at one time. As the graph structure was maintained by the different package (*JGraphT* library), the nodes have no references to other objects except enums, so the only shallow analysis was required.

The package contained 48 different classes and 49096 instances. Further analysis showed that the instances are distributed only among 6 classes of the domain model. The shallow analysis of this namespace took 3.22 hours on the same machine. In this case, the structure of the results was quite different – in each class, multiple triplets of identical instances were discovered.

Further manual analysis of the result showed that the problem, in this case, was in the different sessions. As we went through the triplets, it became obvious that they are part of the same graph – in fact, the duplicates were not only the nodes of the graph, which were discovered by our tool, but the graphs themselves. The reason for this was that two of the users were visualizing the same graph and the server maintained a copy of all the data for both sessions and also – as a form of the cache memory – a third copy not related to any session. As only the visualization was required from the server, it would be possible (in this particular case) to merge all data and maintain only one copy for every user who needs it. This issue is similar to the problems described in [7], as the graphs are also mainly products of the ORM. However, in this case, we cannot really speak about the ORM antipattern, the problem is more in the design of the data structures and lack of sharing data between users in the moments when it is possible.

However, our tool was not able to discover this issue directly, as no package from *JGraphT* was analyzed, so there was no overview of the whole structure during the analysis. This shows obvious limitation when only part of the namespace is analyzed – the data structures that are keeping data are not part of the analysis and even if the deep comparison is used, the identical structures will not be visible.

TABLE III  
ANALYSIS OF CLASSES IN THE ECLIPSE IDE

Package name	Classes	Instances	Found duplicates	Duration [ms]
org	9647	141970	756	5007822
com	919	27906	865	90271
java	1155	313405	39	23596884
sun	929	28092	20	91228
ch	244	539	5	7335

TABLE IV  
ANALYSIS OF CLASSES IN THE INTELLIJ IDEA IDE

Package name	Classes	Instances	Found duplicates	Duration [ms]
org	2016	157743	283	8425230
com	7687	77927	261	1290908
sun	1119	15620	31	26023

Furthermore, for such big structures as the graphs in our case (49096 instances are in fact only data in 5 graphs, without the overhead of the *JGraphT* library), the deep analysis would be quite time demanding – especially if parts of the graph would be shared and only some parts would be changed. On the other hand, this type of the situation – data shared or cloned on the server among several sessions – can be an example when the analysis of the duplicities is useful.

## VI. CONCLUSION AND FUTURE WORK

In the paper, we have presented an algorithm and tool designed to search duplicates in the memory space of the Java applications. The analysis is based on the exploration of heap dump and comparing primitive fields of the objects with the same class. We have implemented both shallow and deep comparisons and demonstrated their functionality on the sample application. According to the results, the algorithm is capable of finding the duplicates that are present in the memory for both simple objects from the testing application and also in the data obtained from four real-world applications. The quadratic complexity of the algorithm, along with the need to compare deep structures, prevents processing a large number of instances, but even on the real heap dumps the algorithm was able to perform the analysis within hours. As this type of the analysis is not something that needs to be performed often, but mainly in the situation when there is a problem with resource consumption, we believe that the tool is practically utilizable.

Main goal of our immediate future work is to implement the parallelization of the task. The shallow comparison should be simply parallelizable, as when all instances from the heap are loaded to the memory of the analyzer, the comparisons need to be performed only within each class and thus can be distributed among the working threads.

As for deep object comparison, the problem is more difficult there, as all instances need to be present in the memory when

a graph of referred objects should be constructed. However, if sufficient memory is available, the task can be still distributed, as each worker can obtain the whole copy of the heap and then work on the analysis of objects within a particular class. The question remains if the communication between such workers would allow making the process faster, for example, if the sub-graphs of the compared objects are already processed and the information about their equality is available. This approach would require to determine a sequence, in which objects should be compared and evaluated, in order to have simpler objects processed before the complex one.

## VII. ACKNOWLEDGMENT

This research was supported by the project LO1506 (PUN-TIS) of the Czech Ministry of Education, Youth and Sports under the program NPU I

## REFERENCES

- [1] N. Mitchell, E. Schonberg, and G. Sevitsky, "Four trends leading to java runtime bloat," *IEEE Software*, vol. 27, no. 1, pp. 56–63, Jan 2010.
- [2] K. Jezek and R. Lipka, "Antipatterns causing memory bloat: A case study," in *2017 IEEE 24th International Conference on Software Analysis, Evolution and Reengineering (SANER)*, Feb 2017, pp. 306–315.
- [3] P. Liden. (2017) String deduplication in gl (accessed: 13 may 2019). [Online]. Available: <http://openjdk.java.net/jeps/192>
- [4] K. Hadj Salem, Y. Kieffer, and S. Mancini, "Formulation and Practical Solution for the Optimization of Memory Accesses in Embedded Vision Systems," in *PROCEEDINGS OF THE 2016 FEDERATED CONFERENCE ON COMPUTER SCIENCE AND INFORMATION SYSTEMS (FEDCSIS)*, ser. ACSIS-Annals of Computer Science and Information Systems, Ganzha, M and Maciaszek, L and Paprzycki, M, Ed., vol. 8, PTI; IEEE. 345 E 47TH ST, NEW YORK, NY 10017 USA: IEEE, 2016, Proceedings Paper, pp. 609–617, Federated Conference on Computer Science and Information Systems (FedCSIS), Gdansk, POLAND, SEP 11-14, 2016.
- [5] G. Xu and A. Rountev, "Precise memory leak detection for java software using container profiling," in *2008 ACM/IEEE 30th International Conference on Software Engineering*, May 2008, pp. 151–160.
- [6] M. Jump and K. S. McKinley, "Cork: Dynamic memory leak detection for garbage-collected languages," *SIGPLAN Not.*, vol. 42, no. 1, pp. 31–38, Jan. 2007. [Online]. Available: <http://doi.acm.org/10.1145/1190215.1190224>
- [7] T.-H. Chen, W. Shang, Z. M. Jiang, A. E. Hassan, M. Nasser, and P. Flora, "Detecting performance anti-patterns for applications developed using object-relational mapping," in *Proceedings of the 36th International Conference on Software Engineering*, ser. ICSE 2014. New York, NY, USA: ACM, 2014, pp. 1001–1012. [Online]. Available: <http://doi.acm.org/10.1145/2568225.2568259>
- [8] N. Mitchell and G. Sevitsky, "The causes of bloat, the limits of health," *SIGPLAN Not.*, vol. 42, no. 10, pp. 245–260, Oct. 2007. [Online]. Available: <http://doi.acm.org/10.1145/1297105.1297046>
- [9] A. E. Chis, N. Mitchell, E. Schonberg, G. Sevitsky, P. O'Sullivan, T. Parsons, and J. Murphy, "Patterns of memory inefficiency," in *Proceedings of the 25th European Conference on Object-oriented Programming*, ser. ECOOP'11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 383–407. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2032497.2032523>
- [10] D. Langr and I. Simecek, "On Memory Footprints of Partitioned Sparse Matrices," in *PROCEEDINGS OF THE 2017 FEDERATED CONFERENCE ON COMPUTER SCIENCE AND INFORMATION SYSTEMS (FEDCSIS)*, ser. Federated Conference on Computer Science and Information Systems, Ganzha, M and Maciaszek, L and Paprzycki, M, Ed., PTI; IEEE. 345 E 47TH ST, NEW YORK, NY 10017 USA: IEEE, 2017, Proceedings Paper, pp. 513–521, Federated Conference on Computer Science and Information Systems (FedCSIS), Prague, CZECH REPUBLIC, SEP 03-06, 2017.
- [11] G. Xu and A. Rountev, "Detecting inefficiently-used containers to avoid bloat," *SIGPLAN Not.*, vol. 45, no. 6, pp. 160–173, Jun. 2010. [Online]. Available: <http://doi.acm.org/10.1145/1809028.1806616>
- [12] O. Shacham, M. Vechev, and E. Yahav, "Chameleon: Adaptive selection of collections," *SIGPLAN Not.*, vol. 44, no. 6, pp. 408–418, Jun. 2009. [Online]. Available: <http://doi.acm.org/10.1145/1543135.1542522>
- [13] A. Infante and A. Bergel, "Object equivalence: Revisiting object equality profiling (an experience report)," *SIGPLAN Not.*, vol. 52, no. 11, pp. 27–38, Oct. 2017. [Online]. Available: <http://doi.acm.org/10.1145/3170472.3133844>
- [14] D. Marinov and R. O'Callahan, "Object equality profiling," in *Proceedings of the 18th Annual ACM SIGPLAN Conference on Object-oriented Programming, Systems, Languages, and Applications*, ser. OOPSLA '03. New York, NY, USA: ACM, 2003, pp. 313–325. [Online]. Available: <http://doi.acm.org/10.1145/949305.949333>
- [15] K. Nasartschuk, M. Dombrowski, K. B. Kent, A. Micic, D. Henshall, and C. Gracie, "String deduplication during garbage collection in virtual machines," in *Proceedings of the 26th Annual International Conference on Computer Science and Software Engineering*, ser. CASCON '16. Riverton, NJ, USA: IBM Corp., 2016, pp. 250–256. [Online]. Available: <http://dl.acm.org/citation.cfm?id=3049877.3049904>
- [16] K. Nasartschuk, K. B. Kent, S. A. MacKay, A. Micic, and C. Gracie, "Improving garbage collection-time string deduplication," in *Proceedings of the 27th Annual International Conference on Computer Science and Software Engineering*, ser. CASCON '17. Riverton, NJ, USA: IBM Corp., 2017, pp. 113–119. [Online]. Available: <http://dl.acm.org/citation.cfm?id=3172795.3172809>
- [17] D. F. Bacon, S. J. Fink, and D. Grove, "Space- and time-efficient implementation of the java object model," in *Proceedings of the 16th European Conference on Object-Oriented Programming*, ser. ECOOP '02. Berlin, Heidelberg: Springer-Verlag, 2002, pp. 111–132. [Online]. Available: <http://dl.acm.org/citation.cfm?id=646159.680023>
- [18] N. Grech, J. Rathke, and B. Fischer, "Jequalitygen: Generating equality and hashing methods," *Sigplan Notices - SIGPLAN*, vol. 46, pp. 177–186, 10 2010.
- [19] E. Aftandilian. (2018) Hprof heap dump parser (accessed: 13 may 2019). [Online]. Available: <https://github.com/eaftan/hprof-parser>
- [20] T. Potuzak and R. Lipka, "Deep object comparison for interface-based regression testing of software components," in *Proceedings of the 2018 Federated Conference on Computer Science and Information Systems, FedCSIS 2018, Poznań, Poland, September 9-12, 2018.*, 2018, pp. 1053–1062. [Online]. Available: <https://doi.org/10.15439/2018F51>
- [21] Oracle. (2019) The jmap utility (accessed: 13 may 2019). [Online]. Available: <https://docs.oracle.com/javase/8/docs/technotes/guides/troubleshoot/tooldescr014.html>
- [22] P. Software. (2019) Spring boot 2.1.4 (accessed: 13 may 2019). [Online]. Available: <https://spring.io/projects/spring-boot>
- [23] I. Eclipse Foundation. (2019) Eclipse ide 2018-12 (accessed: 13 may 2019). [Online]. Available: <https://www.eclipse.org/downloads/>



# Redesigning Method Engineering Education Through a Trinity of Blended Learning Measures

Sietse Overbeek and Sjaak Brinkkemper  
Department of Information and Computing Sciences,  
Faculty of Science, Utrecht University,  
Princetonplein 5, 3584 CC Utrecht, the Netherlands  
Email: {S.J.Overbeek, S.Brinkkemper}@uu.nl

**Abstract**—This paper presents a teaching case of a Blended Learning (BL) approach that was applied to a course on Method Engineering (ME) intended for graduate Business Informatics (BIS) students. The main reason for transforming a Master course on ME from traditional to blended is to take advantage of combining frontal instruction with e-learning based instruction and at the same time reducing lecturers' workload in times of increasing student numbers in BIS and Computer Science (CS) areas. The BL approach consists of three parts, as it consists of the introduction of computer-supported peer assessment, interactive e-lectures, and digital examination. The approach has been reflected upon by course lecturers themselves and it was evaluated through two separate student surveys, from which a variety of positive outcomes can be deduced. Increased generation of feedback, an increase in student motivation, and improved understanding of the course content are three of these outcomes that stand out. On top of student related advantages, especially the BL parts concerning peer assessment and digital examination reduce teaching load. These findings are informative for both education researchers and instructors who are interested in embedding BL in BIS or CS education.

## I. INTRODUCTION

Institutions of higher education are increasingly adopting Blended Learning (BL), the combination of face-to-face and technology-mediated instruction [1, p. 185]. In 2011, scholars noted an “explosive growth of blended learning” and acknowledged its potential to become the “new normal” in higher education [2, pp. 207-208]. In 2017 it was indicated that in the foreseeable future a strengthening will be seen of all kinds of digital learning forms, including various kinds of blended learning [3, p. 216].

The opportunities that BL has on offer provide possible solutions for dealing with current challenges in Business Informatics (BIS) degrees curricula. Prominent examples of such challenges are the preservation of quality education while experiencing increasing student numbers under tight budgets and to provide students an experience that fits their individual learning styles while reducing restrictions on time and place when offering education [4]. This paper explains a teaching case of how BL has been used to redesign a course on Method Engineering (ME) as part of a Master's degree programme in BIS to deal with the aforementioned challenges. The main research question that has guided the redesign process of this ME course is formulated as follows: “How can we provide Method Engineering students a teaching

experience that emphasizes the advantages of blended learning while reducing teaching load at the same time?”.

In section II, the course is outlined and the design of the course is explained before the actual application of BL took place. Section III provides an explanation of the BL approach that is used to realize a scalable and technology-mediated incarnation of the ME course. Students reflected on the BL approach through two surveys and the results of these reflections are found in section IV. Gathered insights and remaining challenges after reflecting on both the experience of running the BL version of the ME course and the student evaluations are discussed in section V. Section VI concludes this paper and gives an overview of future research.

## II. BACKGROUND

The ME course as discussed in this paper is part of a two-year Master's degree programme in BIS. The course is a mandatory course and is offered to students in the first year of the curriculum. Since its inception in 2004, the course has grown from twenty-four to eighty-three participants in 2019. The traditional design of the course consists of regular lectures, lab sessions, and two paper-based exams. During the lab sessions, the students worked individually on their method engineering project. For this project, the students performed a literature review on the topic of a self-selected Information Systems (IS) development method or technique and based on the gathered data from the literature study they designed a meta-model of the selected method or technique. This project was split in five parts and each of the parts was completed by providing a deliverable of which the final deliverable consisted of a term paper that is an integration of the previous deliverables.

The year 2017 was the final year when the ME course ran in a traditional way. With seventy-one participants producing a total of five deliverables and two written exams within a course period of eleven weeks there was substantial pressure on the teaching team to provide feedback on the student materials and grade them. The team consisted of two lecturers and three student assistants (SAs), sometimes also called teaching assistants. The SAs provided guidance during the lab sessions and provided practical support. With eighty-three participants signed up for the 2018 incarnation of the course, the time was ripe to implement proper changes in the design

of the course in order to deal with increasing student numbers without sacrificing the quality of the course while reducing the teaching load for the entire team. In the following section, the different measures in redesigning the course are further explained.

### III. METHOD ENGINEERING EDUCATION MADE BLENDED AND SCALABLE

From a birds-eye perspective, the application of BL in the ME course consists of three overall measures, which are: 1) the implementation of computer-supported peer assessment, 2) the introduction of interactive e-lectures, and 3) introducing digital exams. Each of these three overall measures are described in detail in the following sub sections.

#### A. Computer-supported peer assessment

The tool 'Revisely' (see: <https://revise.ly>) is used to cater for online submission of all different deliverables as part of the ME project and, more importantly, it allows for the introduction of peer assessment in the form of peer reviewing and peer grading. By introducing peer assessment, students are able to learn from each other's work. As the students have to work on their own topic for the meta-modelling project, they are able to acquire useful insights when reviewing the modelling choices made by fellow students. Moreover, the students gain experience in providing feedback and grading the work of peers. Finally, through peer assessment a student not only receives feedback from lecturers or student assistants but also from their peers [5, p. 132]. A total of three peer assessment exercises were introduced in the ME project. For the first exercise, students have to peer review and grade a deliverable in which the selected IS development method or technique is explained and positioned relative to existing literature. For the second exercise students have to peer review and grade a deliverable that includes the design of a meta-model of the selected method or technique and for the third exercise students peer reviewed the pre-final term paper without grading it. Based on the feedback acquired from their peer assessors, students write their final term paper that is subsequently assessed and graded by the lecturing team itself. Figure 1 shows a screenshot of the Revisely tool that is used for computer-mediating the peer assessment exercises. Students uploaded their deliverables as PDF documents in the Revisely tool, where a total of three randomly assigned peer assessors needed to provide both textual feedback, i.e., remarks and suggestions for improvement, and scores for the individual grading criteria. These criteria are made available in the Revisely environment and are in fact based on the grading criteria as used by the lecturing team itself in the previous year. To provide the peer assessors with a frame of reference, they are provided with examples from the previous year that were assessed as mediocre / weak, sufficient, and excellent by lecturers. To make sure the peer assessment exercises are conducted in a serious manner, they are required elements for course completion as they are part of the exam rules of the course. This has as a consequence that failure to

deliver a serious peer review leads to an inability to pass the course, while the peer assessment results, i.e., the quality of the peer feedback is not graded by the lecturers. For each peer assessment exercise, we found that almost everybody submitted a serious peer assessment. Both the fact that the peer assessments are required elements to pass the course and that three students peer assess each deliverable may further stimulate students to take this task seriously [5, p. 103].

After the passing of a peer assessment deadline, the student assistants are asked to inspect the peer assessments, i.e., the feedback and grades given by the peer assessors. In the Revisely tool, every student assistant is assigned an equal share of the deliverables that are indeed commented and graded by three students. A student assistant then checks if the grades that are provided are fair and in line with the quality of a deliverable, if there are outliers in the provided grades, and if the peer review is conducted according to the grading criteria as made available in Revisely. A student assistant also extends provided feedback if needed, or adds additional feedback if additions to the peer reviews are needed. As a final step, a student assistant proposes a final grade based on the grades as provided by means of the peer grading activities. When the student assistants are all done performing this 'meta review' of the provided peer reviews, the lecturers conduct a final 'meta meta review', i.e., discussing conspicuities as identified by the student assistants, performing a final check of the meta reviews, and determining the final grade based on the proposals made by the student assistants.

At the end of the day the utilization of this computer-supported approach to peer reviewing and peer grading provided at least six clearly identifiable advantages. First and foremost the lecturing team including the student assistants experienced a relieved teaching load as three project deliverables do not need to be reviewed and graded 'from scratch' by lecturers with assistance from student assistants. Secondly, in the past a student would receive feedback on a deliverable from a lecturer and a student assistant. With the peer assessment procedure a student now receives feedback from three fellow students, a student assistant, and possibly a lecturer. Before a lecturer determines the final grade, in fact four suggestions for such a grade are now made by those who have inspected the deliverable. Thirdly, the deliverable submission procedure is streamlined and automated because of the usage of Revisely. Grading criteria are provided online, a randomized match is made between the peer assesseees and assessors, student assistants are allocated a fair share of the assessed deliverables and all feedback and grades are made available online for every individual student. Fourthly, the university where this blended Method Engineering course is offered has a support team for lecturers who incorporate BL in their courses, which means that whenever a user of a BL tool that is supported by the university has a tool-related question the support team can step in and there is no need to communicate with the tool supplier itself. Fifthly, through the above approach students are able to experience an increased level of responsibility [6, p. 88], i.e., it is their task to not only run a successful ME project

The screenshot displays the Revisely interface. On the left is a dark blue sidebar with navigation options: Dashboard, Assignments, My students, My groups/classes, Management, Help, and Logout. The main content area shows two assignment cards. The top card, 'Assignment C: Meta-modeling with PDDs', is dated 2018-03-02 18:00 to 2018-03-17 10:00. It shows a score of 77/77 (Handed in) and 0 (To be corrected), with an average score of 7,27. The bottom card, 'Assignment A: Topic selection and description', is dated 2018-02-12 10:00 to 2018-03-09 22:00. It shows a score of 77/77 (Handed in) and 0 (To be corrected), with an average score of 7,21. Both cards include download options for reports and peer grading links.

Fig. 1. Screenshot of the Revisely tool

but also to conduct peer assessments in a serious manner. This way, they gain experience in giving feedback to peers and in grading each other's work. Fifthly, lecturers are able to identify those students who come up with high quality reviews, which is a factor in identifying whether they are potential candidates to become the students assistants of the future [7]. Sixthly and finally, introducing peer assessment exercises in addition to the tasks to deliver the ME project deliverables is a way to increase student engagement during the course, which is helpful in times where it is easy for students to spend time on non-study related activities [5, p. 132].

### B. Interactive e-lectures

Introduction of interactive e-lectures is the second of the three overall measures taken to apply BL in the ME course as discussed in this paper. A gradual approach is adopted to modify selected regular offline lectures into e-lectures. For the 2018 incarnation of the course, three of the in total eleven regular lectures are transformed into e-lectures. These three lectures are foundational lectures and discuss the topics of meta-data modelling, meta-process modelling, and the role of formalization in ME. Gaining experience in this new teaching mode first before modifying the other regular lectures that discuss advanced ME topics is the key reason for applying such a gradual approach. To modify regular lectures into e-lectures, the tool 'Scalable Learning' (see: [www.scalable-learning.com](http://www.scalable-learning.com)) is used and just as with Revisely this tool has an university-based support team. The e-lectures themselves consist of video material from previously recorded lectures and knowledge clips that were recorded by a student assistant. This student assistant was specifically appointed for this task and recorded short topical clips in a studio on the university campus. These topical clips are in line with what is taught in the three regular lectures. For each e-lecture in the Scalable Learning tool, multiple-choice quiz questions are

added to make the e-lectures interactive. The students are asked to prepare each e-lecture, i.e., watch them and make the questions at home or at whatever place they wish. In the lecturer view in Scalable Learning, it is then possible to see who has completed an e-lecture and the answers to the quiz questions can be inspected. The lecturer can then make a selection which questions need to be discussed offline in a classroom setting, for example, those questions that were difficult to answer correctly for the students. Students are also allowed to add their own comments or questions to the video material if there are unclarities while watching the e-lectures. These inaudibilities can also be highlighted for discussion in class. After the students finalized preparation of an e-lecture, an in-class discussion session followed of about an hour each where quiz questions that proved to be difficult and any other unclarities related to the online course material are discussed.

Clear advantages of the approach as described above are threefold. Firstly, students are able to watch the e-lectures anywhere and on their own learning pace, i.e., quickly going over those clips and questions that are easier to grasp for an individual student and more slowly watching material or even repeatedly watching material that is more difficult. Secondly, time is gained as the necessity to be physically present in lecture rooms is reduced. Thirdly, for the overall population of students the lecturer can identify which questions are particularly difficult or more easy to answer and by reviewing the quiz questions the lecturer can shape the offline discussion sessions in order to attune these to the specific audience.

### C. Digital mid-term and final exams

The third and final measure taken to realize a blended variant of the ME course is the transformation from paper-based exams to digital exams. This measure together with the electronic submission of the ME project deliverables in Revisely has as additional side effect that the ME course is

now entirely paperless. The university-supported tool Remindo (see: [www.paragin.nl](http://www.paragin.nl)) is used to design the digital exams. In previous years, the ME course included two paper-based exams, i.e., a mid-term exam and a final exam. The mid-term exam is meant to test knowledge on meta-modelling gained in the first half of the course, while the final exam is for the larger part meant to test knowledge on method engineering theories. The paper-based exams included open questions, while Remindo is ideally suited for administering multiple-choice exams. Introducing multiple-choice questions in both exams provides an opportunity to further reduce the teaching load, as no manual marking is needed for those question types. Transforming open questions where students are asked to design (meta-)models requires to be notably creative, e.g., such an open question can be transformed to a closed variant by dividing the open question in parts where for each closed question a student has to choose the correct modelling alternative from a set of choices [8, p. 464]. Another option is to show a partial (meta-)model that a student has to complete by correctly dragging-and-dropping modelling elements such as correctly dragging-and-dropping a meta-activity in the eventual (meta-)model or correctly positioning meta-concepts in the eventual (meta-)model.

The most prominent advantage of this approach is that the time needed for marking is reduced. Instead of grading two times eighty-three exams, Remindo takes most grading work out off the hands as it automatically checks closed questions and is able to deal with negate guessing. Needless to say, digital exams save time and paper as printing of big stacks of exams and carrying these to and from exam rooms becomes a thing of the past. Finally, in case of open questions Remindo removes the possibility of having to deal with unreadable handwriting, as students have to type their answers on notebooks that run Remindo.

#### IV. STUDENTS' REFLECTIONS ON THE BLENDED LEARNING APPROACH

The trinity of BL measures as applied in the 2018 incarnation of the discussed ME course has been evaluated by students through: 1) A customary online survey purposefully tailored with specific BL-related open questions that is presented to students at the end of all courses that are part of the Master's degree programme in BIS of which also the mentioned ME course is part of, and 2) through a survey offered by the university's BL support team. The latter survey measures student motivation, differences experienced in learning activities, and experienced learning outputs related to the computer-supported peer assessment part of the trinity [6], [9] and was also offered to students at the end of the course.

Inspired by Unkelos-Shpigel and Hadar [10, p. 189], the survey data has been analyzed in an inductive manner [11], [12] with respect to the part of the main research question that concerns the students' teaching experience by boosting the advantages of blended learning. The customary online survey shows insights on experiences of all three BL measures and it was filled out by thirty-eight of the eighty-three participants in

2018. The following responses on peer assessment are found and they happen to be rather self-explanatory:

*"I like Revisely and did not have any issues with it. Grading others' work gives better insight into your own work"*

*"I would keep the assignment format and Revisely. It is nice how each assignment builds into the final paper. It was really stress-free and I enjoyed it"*

*"Revisely provides an OK platform for submissions and peer reviews"*

The responses related to the interactive e-lectures show that students perceived an added value in the activating effect these kind of lectures have and in the 'blended' aspect of having online and offline lectures. The apparent usefulness of the e-lectures is also emphasized:

*"The e-lectures [...] demanded active participation [by means of answering] the [quiz] questions"*

*"Enjoyed the different elements in the course, [I] like the combination of the regular lectures with the e-lectures"*

*"e-Lectures are a great addition in my opinion"*

*"The e-lectures should [remain for next year, they were] very useful"*

The following responses concern the third and final BL measure which is the introduction of digital exams:

*"I found the flipped classroom [and] digital exams [...] very helpful"*

*"The mid-term exam was good. I'm a big fan of the digital exam and it was carefully constructed as to test the knowledge of [meta-modelling] despite the [common view] that open exams would be a better way to [test this kind of knowledge]"*

*"While I'm not a fan of multiple choice questions, I'm glad the majority of the questions were about understanding the content as opposed to [sheer memorizing of] it"*

One response concerned a positive impression of the course as a whole:

*"Compared to last year the course has really improved"*

This impression resonates when comparing last year's average grade given for the course by students in 2017 with this year's average grade, as the respondents in 2018 evaluated the overall quality of this course to be a 7.5 on a scale of 1 to 10 (N=32 with a standard deviation of 1), whereas last year this score was a 6.5 (N=26, standard deviation of 1.5).

The survey offered by the university's BL support team was filled out by a total of seventy-one out of the eighty-three course participants in 2018. The ME students provided their answers on a five-point Likert scale, where '1' means 'completely disagree' and '5' means 'completely agree'. In table I the averages are shown for the ME course and for two other courses that also used Revisely for computer-supported peer assessment with a total of one hundred and

fourteen respondents. In the rightmost column the averages are shown for all courses within the university that are using university-supported BL tooling since 2016. In that year the university's BL support team was formed and since then surveys are offered to those lecturers who use university-supported BL tools. The table shows that students indicated to have received more feedback and also gave more feedback by using computer-supported peer assessment in the ME course. This is an indication that the main aim of the peer assessment approach is largely met, which is to offer an environment to provide peer feedback and peer grading. On average, students were also highly positive about the way the peer assessment approach affected their motivation and that they were able to learn more and understand the course content better. The students were less positive about 1) the joy experienced while using computer-supported peer assessment, 2) the effect it had on their ability to pass the exams, and 3) the extent to which it helped them to understand the lectures better. In the future, it will be investigated how to improve on these three less positive aspects.

## V. DISCUSSION

After analyzing the results of the student surveys in full, there are other observations that are deemed relevant for further discussion apart from the three more critical aspects as mentioned at the end of section IV. After discussing these observations, some interesting points that are specifically related to the peer assessment part are mentioned at the end of this section.

### Meta-review to overcome peer reviews of varying quality

Concerning the peer assessments, respondents indicated that the quality of peer reviews may differ and grades given for peer graded deliverables are not always in line with each other. These effects were anticipated on in the design of the blended variant of the ME course by introducing three peer reviewers [13, p. 43] and by conducting a round of meta- and meta meta reviewing as discussed in section III-A. However, it does not prevent students from experiencing differences in quality and differences in grades given to their work as the results of a peer assessment are visible in the Revisely tool once a peer assessment is finalized by a fellow student. Although the final grade for their work is given by a lecturer, a respondent wondered what effect an outlier had on the final grade as this was not consistently articulated in the final feedback. Apart from an obvious solution to make explicit in the final feedback of the meta-meta-review what has been done with a possible outlier when determining a final grade, the introduction of an entire instructional lecture dedicated to peer assessment would be a plausible idea, instead of maintaining the current practice of instructing students in an ad hoc manner how to conduct a peer assessment as part of a topical lecture and during workshops where they work on their practical exercises [9, p. 103].

### Student opinions in two camps

In the responses concerning the interactive e-lectures it was found not every student liked the idea of having to watch the e-lectures and prepare the accompanying quiz questions. Some students feel a lack of opportunity to interact with others, i.e., lecturer and fellow classmates in a live classroom setting. However, the Scalable Learning tool offers the possibility for lecturers to comment on questions that are raised by students while preparing the e-lectures and to at least asynchronously interact with students before the actual in-class discussion session takes place. What will be done for the next iteration of the course is to explicitly emphasize in class the possibilities to asynchronously communicate with lecturers by making use of the option to raise questions and clearly indicate where unclarities are. Watching interactive e-lectures remains a different form of education when compared to a traditional lecture, however, and it depends on an individual's learning preferences how this modern educational form is experienced. Another notable aspect which is in fact sensible advice is that students indicate the e-lectures should always include knowledge clips that are of identical quality when compared to regular lectures and that lecturers should prevent knowledge clips from becoming second-rate replacements of regular lectures.

### Multiple-choice exams with an option to comment

From the responses to the customary online survey it becomes clear that among students who are unfamiliar with multiple-choice exams or who have a preference for open question exams there is a desire to provide comments in the digital exam environment next to only being able to select the proper answers. Multiple-choice exams with an option to comment provides students with an opportunity to write down their thought-line that led to the selection of an answer. Nield and Wintre [14] indicated that this approach reduces frustration and produces less anxiety among students. How to implement this in the digital exam environment is a different matter, i.e., it is not a standard option to choose from in the Remindo tool and as such how to deal with this challenge is part of future research.

### Overcoming glitches in technology-mediated peer assessment

Concerning the technology-mediated peer assessment part there are some noteworthy experiences from the point of view of the ME course lecturers. As mentioned before, after the submission deadline of a deliverable that was going to be peer reviewed in Revisely had passed, three reviewers were assigned to each submission. It was found that the tool supported random assignment of one reviewer only, meaning the assignment of three reviewers had to be done manually which is a more time consuming and error-prone process. The tool also

TABLE I  
COMPUTER-SUPPORTED PEER ASSESSMENT EVALUATION RESULTS

By using the computer-supported peer assessment approach I ... :	ME course <i>N=71</i>	Other courses <i>N=114</i>	Totals for BL courses <i>N=3135</i>
<i>Motivation</i>			
was motivated	3.69	3.49	3.46
experienced joy	2.69	2.91	3.33
had the impression it was useful	4.03	3.86	3.73
had the impression it supported me in passing the exams	2.92	2.99	3.19
<i>Learning activities</i>			
was more active with the content	3.33	3.02	3.47
was able to better study the content	3.32	2.95	3.17
was able to improve collaboration	3.00	2.91	2.72
received more feedback	4.37	4.27	3.32
gave more feedback	4.28	3.70	3.13
<i>Experienced learning outputs</i>			
learned more in this course	3.54	3.28	3.32
understood the content better	3.44	3.17	3.30
was better prepared for the exams	2.79	2.74	3.16
was better prepared for the lectures	2.97	2.85	2.84
understood the lectures better	2.69	2.53	3.01

has the functionality that if students finalize their peer review, the reviewee can immediately see the feedback. This has as an advantage that feedback is seen much quicker when compared to an instructor review approach. However, it was found that once a peer reviewer marks the review as finalized, there is no option to undo. It happened a couple of times that a peer reviewer would mark a review as finalized accidentally. As a workaround additional feedback could then be exchanged between the reviewer and reviewee by e-mail. Finally, a pressing issue was that three students were not allowed to review the same deliverable at the same time. If they did, added feedback could be lost. As a workaround students were asked to store their review on a local drive first and then they were asked to always communicate with the other peer reviewers of the deliverable in case one of the three reviewers would start performing a review. The above experiences were all communicated to the university's BL support team who maintained close ties with the tool supplier and as such formed improvement points for the next release of the tool.

## VI. CONCLUSIONS AND FUTURE RESEARCH

In this paper a redesign approach was presented for transforming a Master's course on Method Engineering as part of an BIS curriculum into a blended variant. A trinity of blended learning measures have been proposed, that were driven by the desire to realize 1) a course design that preserves quality education in a time where student numbers in CS and BIS curricula are increasing while university budgets remain tight, and 2) to offer a teaching experience that fits with individual learning styles while reducing time and place restrictions. The three measures included the introduction of technology-mediated peer assessment, interactive e-lectures, and digital examination. Most prominent advantages of the peer assess-

ment measure are the reduction of teaching load, increased generation of feedback itself, increased student experience in providing feedback, and increased student motivation and engagement. The introduction of interactive e-lectures enabled students to watch the e-lectures anywhere, anytime, and on their own learning pace. There is a time gain as e-lectures replace regular lectures, however, quiz results need to be reviewed by lecturers and then in-class discussion sessions need to be organized. This offers the advantage of tailoring these sessions in such way that the difficult questions as part of the e-lectures receive most attention. Clear benefits of digital examination include reduced marking time and elimination of paper-based exams.

For next year, the blended approach will be maintained in the Method Engineering course and improvements will be implemented, based on opportunities identified in the student survey results and by means of reflecting on the course from a lecturer's point of view. An instructional lecture purely dedicated to peer assessment will be introduced, possibilities to stimulate student-student and student-lecturer interaction will be explored when students are watching e-lectures, e-lecture quality will be double-checked and where needed further improved, and it will be investigated whether multiple-choice exams with an option to comment can be realized. Finally, based on the peer assessment survey results it will be investigated how students are able to experience more joy while using computer-supported peer assessment (and BL tools in general for that matter), in what ways peer assessment can make an impact on the ability to pass exams, and it will be inventorized how it helps students to increase understanding of the course content.

## REFERENCES

- [1] W. W. Porter, C. R. Graham, K. A. Spring, and K. R. Welch, "Blended learning in higher education: Institutional adoption and implementation,"

- Computers & Education*, vol. 75, pp. 185–195, 2014, doi: [10.1016/j.compedu.2014.02.011](https://doi.org/10.1016/j.compedu.2014.02.011).
- [2] A. Norberg, C. D. Dziuban, and P. D. Moskal, “A timebased blended learning model,” *On the Horizon*, vol. 19, no. 3, pp. 207–216, 2011, doi: [10.1108/10748121111163913](https://doi.org/10.1108/10748121111163913).
- [3] B. van der Zwaan, *Higher Education in 2040: A Global Approach*. Amsterdam, the Netherlands: Amsterdam University Press, 2017.
- [4] D. Garrison and H. Kanuka, “Blended learning: Uncovering its transformative potential in higher education,” *The Internet and Higher Education*, vol. 7, no. 2, pp. 95–105, 2004, doi: [10.1016/j.iheduc.2004.02.001](https://doi.org/10.1016/j.iheduc.2004.02.001).
- [5] B. Jones, *Motivating Students by Design: Practical Strategies for Professors*. Blacksburg, USA: Brett D. Jones, 2018.
- [6] R. M. Filius, R. A. de Kleijn, S. G. Uijl, F. J. Prins, H. V. Rijen, and D. E. Grobbee, “Strengthening dialogic peer feedback aiming for deep learning in SPOCs,” *Computers & Education*, vol. 125, pp. 86–100, 2018, doi: [10.1016/j.compedu.2018.06.004](https://doi.org/10.1016/j.compedu.2018.06.004).
- [7] D. G. Collings, H. Scullion, and P. M. Caligiuri, Eds., *Global Talent Management*, 2nd ed. New York, USA: Routledge, 2019.
- [8] A. De Lucia, C. Gravino, R. Oliveto, and G. Tortora, “An experimental comparison of ER and UML class diagrams for data modelling,” *Empirical Software Engineering*, vol. 15, no. 5, pp. 455–492, 2010, doi: [10.1007/s10664-009-9127-7](https://doi.org/10.1007/s10664-009-9127-7).
- [9] R. M. Filius, R. A. de Kleijn, S. G. Uijl, F. J. Prins, H. V. Rijen, and D. E. Grobbee, “Promoting deep learning through online feedback in SPOCs,” *Frontline Learning Research*, vol. 6, no. 2, pp. 92–113, 2018, doi: [10.14786/flr.v6i2.350](https://doi.org/10.14786/flr.v6i2.350).
- [10] N. Unkelos-Shpigel and I. Hadar, “Test first, code later: Educating for test driven development,” in *Advanced Information Systems Engineering Workshops*, R. Matulevičius and R. Dijkman, Eds. Cham, Switzerland: Springer International Publishing, 2018, pp. 186–192, doi: [10.1007/978-3-319-92898-2\\_16](https://doi.org/10.1007/978-3-319-92898-2_16).
- [11] B. Oates, *Researching Information Systems and Computing*. Thousand Oaks, USA: SAGE Publications, 2006.
- [12] A. Strauss and J. Corbin, *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Thousand Oaks, USA: SAGE Publications, 1998.
- [13] S. P. Balfour, “Assessing writing in MOOCs: Automated essay scoring and calibrated peer review,” *Research & Practice in Assessment*, vol. 8, pp. 40–48, 2013.
- [14] A. F. Nield and M. G. Wintre, “Multiple-choice questions with an option to comment: Student attitudes and use,” *Teaching of Psychology*, vol. 13, no. 4, pp. 196–199, 1986, doi: [10.1207/s15328023top1304\\_6](https://doi.org/10.1207/s15328023top1304_6).



# 3<sup>rd</sup> International Conference on Lean and Agile Software Development

THE evolution of software development life cycles is driven by the perennial quest on how to organize projects for better productivity and better quality. The traditional software development projects, which followed well-defined plans and detailed documentations, were unable to meet the dynamism, unpredictability and changing conditions that characterize rapidly changing business environment. Agile methods overcame these limits by considering that requirements are not static but dynamic, while customers are unable to definitively state their needs up front. However, the advent of agile methods divided the software engineering community into opposing camps of traditionalists and agilists. After more than a decade of debate and experimental studies a majority consensus has emerged that each method has its strengths as well as limitations, and is appropriate for specific types of projects, while numerous organizations have evolved toward the best balance of agile and plan-driven methods that fits their situation.

In more recent years, the software industry has started to look at lean software development as a new approach that could complement agile methods. Lean development further expands agile software development by adopting practices from lean manufacturing. Lean emphasizes waste elimination by removing all nonvalue-adding activities.

## TOPICS

The objective of LASD is to extend the state-of-the-art in lean and agile software development by providing a platform at which industry practitioners and academic researchers can meet and learn from each other. We are interested in high quality submissions from both industry and academia on all topics related to lean and agile software development. These include, but are not limited to:

- Combining lean and agile methods for software development
- Lean and agile requirements engineering
- Scaling agile methods
- Distributed agile software development
- Challenges of migrating to lean and agile methods
- Balancing agility and discipline
- Agile development for safety systems
- Lean and agility at the enterprise level
- Conflicts in agile teams
- Lean and agile project production and management
- Collaborative games in software processes
- Lean and agile coaching
- Managing knowledge for agility and collaboration

- Tools and techniques for lean and agile development
- Measurement and metrics for agile projects, agile processes, and agile teams
- Innovation and creativity in software engineering
- Variability across the software life cycle
- Industrial experiments, case studies, and experience reports related to all of the above topics
- Gamification
- Affective Software Engineering

## EVENT CHAIRS

- **Przybyłek, Adam**, Gdansk University of Technology, Poland

## PROGRAM COMMITTEE

- **Ahmad, Muhammad Ovais**, University of Oulu, Finland
- **Akman, Ibrahim**, Atılım University, Turkey
- **Ali, Sikandar**, China University of Petroleum, China
- **Almeida, Fernando**, University of Porto & INESC TEC, Portugal
- **Alshayeb, Mohammad**, King Fahd University of Petroleum and Minerals, Saudi Arabia
- **Angelov, Samuil**, Fontys University of Applied Sciences, The Netherlands
- **Bach-Dąbrowska, Irena**, WSB Gdańsk, Poland
- **Bagnato, Alessandra**, SOFTEAM R&D Department, France
- **Belle, Alvine Boaye**, École de Technologie Supérieure, Canada
- **Benayed, Nourchene**, Higher Colleges of Technology, United Arab Emirates
- **Bernhart, Mario**, Vienna University of Technology, Austria
- **Bhadauria, Vikram**, Texas A&M International University, United States
- **Binti Abdullah, Nik Nailah**, Monash University Malaysia, Malaysia
- **Biró, Miklós**, Software Competence Center Hagenberg and Johannes Kepler University Linz, Austria
- **Blech, Jan Olaf**, RMIT University, Australia
- **Borg, Markus**, SICS Swedish ICT AB, Sweden
- **Brzeski, Adam**, CTA.ai, Gdansk University of Technology
- **Buchalceva, Alena**, University of Economics, Prague, Czech Republic
- **Buchan, Jim**, Auckland University of Technology, New Zealand

- **Buglione, Luigi**, Engineering Ingegneria Informatica SpA, Italy
- **Chatzigeorgiou, Alexandros**, University of Macedonia, Greece
- **Cruzes, Daniela**, SINTEF ICT, Norway
- **Daszczuk, Wiktor Bohdan**, Warsaw University of Technology, Poland
- **Dejanović, Igor**, Faculty of Technical Sciences, Novi Sad, Serbia
- **Derezińska, Anna**, Warsaw University of Technology, Institute of Computer Science, Poland
- **Diebold, Philipp**, Fraunhofer IESE, Germany
- **DUTTA, ARPITA**, NIT ROURKELA, India
- **Escalona, Maria Jose**, Universidad de Sevilla, Spain
- **Essebaa, Imane**, Hassan II University of Casablanca, Morocco
- **Fagerholm, Fabian**, University of Helsinki, Finland and Blekinge Institute of Technology, Finland
- **Figueira Filho, Fernando Marques**, Universidade Federal do Rio Grande do Norte, Brazil
- **García-Mireles, Gabriel Alberto**, Universidad de Sonora, Mexico
- **Ghofrani, Javad**, HTW Dresden University of Applied Sciences, Germany
- **Goczyła, Krzysztof**, Gdańsk University of Technology, Poland
- **GODBOLEY, SANGHARATNA**, NIT ROURKELA, India
- **Gonzalez Huerta, Javier**, Blekinge Institute of Technology, Sweden
- **Górski, Janusz**, Gdańsk University of Technology, Poland
- **Gregory, Peggy**, University of Central Lancashire, United Kingdom
- **Hanslo, Ridwaan**, Council for Scientific and Industrial Research, South Africa
- **Heil, Sebastian**, Chemnitz University of Technology, Germany
- **Hohenstein, Uwe**, Siemens AG, Germany
- **Hohl, Philipp**, ZF Friedrichshafen AG, Germany
- **Ikonen, Marko**, Projektivarikko Oy, Finland
- **Janes, Andrea**, Free University of Bolzano, Italy
- **Järvinen, Janne**, F-Secure Corporation, Finland
- **Jarzębowicz, Aleksander**, Gdansk University of Technology, Poland
- **Jovanović, Miloš**, University of Novi Sad, Serbia
- **Kakarontzas, George**, Aristotle University of Thessaloniki, Greece
- **Kaloyanova, Kalinka**, Sofia University, Bulgaria
- **Kanagwa, Benjamin**, Makerere University, Uganda
- **Kapitsaki, Georgia**, University of Cyprus, Cyprus
- **Karolyi, Matěj**, Institute of Biostatistics and Analyses, Faculty of Medicine, Masaryk University, Brno, Czech Republic
- **Karpus, Aleksandra**, Gdańsk University of Technology, Poland
- **Kassab, Mohamad**, Innopolis University, Russia
- **Katić, Marija**, School of Computing, Engineering and Physical Sciences, United Kingdom
- **Khelif, Wiem**, University of Sfax, Tunisia
- **Knodel, Jens**, Fraunhofer IESE, Germany
- **Kropp, Martin**, University of Applied Sciences and Arts Northwestern Switzerland, Switzerland
- **Kuvaja, Pasi**, University of Oulu, Finland
- **Laanti, Maarit**, Nitor, Finland
- **Lehtinen, Timo O. A.**, Aalto University, Finland
- **Lenarduzzi, Valentina**, Tampere University, Finland
- **Liebel, Grischa**, University of Gothenburg, Sweden
- **Luković, Ivan**, University of Novi Sad, Serbia
- **Lunesu, Iaria**, Università degli Studi di Cagliari, Italy
- **Mahnič, Viljan**, University of Ljubljana, Slovenia
- **Mangalaraj, George**, Western Illinois University, United States
- **Marcinkowski, Bartosz**, Department of Business Informatics, University of Gdansk, Poland
- **Mazzara, Manuel**, Innopolis University, Russia
- **Mesquida Calafat, Antoni-Lluís**, University of the Balearic Islands, Spain
- **Miler, Jakub**, Faculty of Electronics, Telecommunications And Informatics, Gdansk University of Technology, Poland, Poland
- **miller, gloria**, Skema Business School, Germany
- **Misra, Sanjay**, Covenant University, Nigeria
- **Mohapatra, Durga Prasad**, NIT ROURKELA, India
- **Morales Trujillo, Miguel Ehecattl**, University of Canterbury, New Zealand
- **Mordinyi, Richard**, Vienna University of Technology, Austria
- **MüncH, Jürgen**, Reutlingen University, Germany
- **Muñoz, Mirna**, Centro de Investigación en Matemáticas, Mexico
- **Muszyńska, Karolina**, University of Szczecin, Poland
- **Nguyen-Duc, Anh**, University College of Southeast Norway, Norway
- **Noyer, Arne**, University of Osnabrueck and Willert Software Tools GmbH, Germany
- **Oktaba, Hanna**, National Autonomous University of Mexico, Mexico
- **Ortu, Marco**, University of Cagliari, Italy
- **Oyetoyan, Tosin Daniel**, SINTEF Digital, Norway
- **Özkan, Necmettin**, Kuveyt Turk Participation Bank, Turkey
- **Panda, Subhrakanta**, Birla Institute of Technology and Science, Pilani, India
- **Pereira, Rui Humberto R.**, Instituto Politecnico do Porto - ISCAP, Portugal
- **Poniszewska-Maranda, Aneta**, Institute of Information Technology, Lodz University of Technology, Poland
- **Poth, Alexander**, Volkswagen AG, Germany
- **Przybyłek, Michał**, University of Warsaw, Poland
- **Ramsin, Raman**, Sharif University of Technology, Iran
- **Ristić, Sonja**, University of Novi Sad, Faculty of Tech-

nical Sciences, Serbia

- **Rossi, Bruno**, Masaryk University, Czech Republic
- **Rybola, Zdenek**, FIT CTU in Prague, Czech Republic
- **Salah, Dina**, Sadat Academy, Egypt
- **Salnitri, Mattia**, University of Trento, Italy
- **Schön, Eva-Maria**, University of Seville, Spain
- **Sedeno, Jorge**, University of Seville, Spain
- **Senapathi, Mali**, Auckland University of Technology, New Zealand
- **Sikorski, Marcin**, PJWSTK
- **Śmiałek, Michał**, Politechnika Warszawska, Poland
- **Soares, Michel**, Federal University of Sergipe, Brazil
- **Soria, Álvaro**, ISISTAN Research Institute, Argentina
- **Spichkova, Maria**, RMIT University, Australia
- **Springer, Olga**, Gdańsk University of Technology, Poland
- **Stålhane, Tor**, Norwegian University of Science and Technology, Norway
- **Stettina, Christoph Johann**, Leiden University, The Netherlands
- **Taibi, Davide**, Free University of Bolzano, Italy
- **Tarhan, Ayca**, Hacettepe University Computer Engineering Department, Turkey
- **Thomaschewski, Jörg**, University of Applied Sciences Emden/Leer, Germany
- **Torrecilla Salinas, Carlos**, University of Seville, Spain
- **Unterkalmsteiner, Michael**, Blekinge Institute of Technology, Sweden
- **Wardziński, Andrzej**, Gdańsk University of Technology, Poland
- **Werewka, Jan**, AGH University of Sci. and Technology, Poland
- **Winter, Dominique**, University of Applied Sciences Emden/Leer, Germany
- **Wróbel, Michał**, Gdańsk University of Technology, Poland
- **Yilmaz, Murat**, Çankaya University, Turkey
- **Zarour, Nacer Eddine**, University Constantine2, Algeria
- **Łukasiewicz, Katarzyna**, Gdańsk University of Technology, Poland



# Preliminary Citation and Topic Analysis of International Conference on Agile Software Development Papers (2002-2018)

Muhammad Ovais Ahmad

Faculty of Electronics, Telecommunications and Informatics,  
Gdansk University of Technology, Poland.

M3S Research Unit, University of Oulu, Finland.

Department of Mathematics and Computer Science,

Karlstad University, Sweden.

Email: ovais.ahmad@oulu.fi

Päivi Raulamo-Jurvanen

M3S Research Unit,

University of Oulu, Finland.

Email: Paivi.Raulamo-Jurvanen@oulu.fi

**Abstract**—This study utilizes citation analysis and automated topic analysis of papers published in International Conference on Agile Software Development (XP) from 2002 to 2018. We collected data from Scopus database, finding 789 XP papers. We performed topic and trend analysis with R/RStudio utilizing the text mining approach, and used MS Excel for the quantitative analysis of the data. The results show that the first five years of XP conference cover nearly 40% of papers published until now and almost 62% of the XP papers are cited at least once. Mining of XP conference paper titles and abstracts result in these hot research topics: “Coordination”, “Technical Debt”, “Teamwork”, “Startups” and “Agile Practices”, thus strongly focusing on practical issues. The results also highlight the most influential researchers and institutions. The approach applied in this study can be extended to other software engineering venues and applied to large-scale studies.

## I. INTRODUCTION

In every field of science, evidence for the importance of identifying emerging research topics is useful for researchers, funding agencies and policy makers. This helps to promote and enhance the development of potentially promising research topics. Citation is a way to judge influential work and build new studies on existing research results [1], [2], [3]. Citation analysis is a common way not only to judge but also to observe the most popular and influential work [1], [2], [4]. Bibliometrics, on the other hand, is a method used for statistical analysis of publications in order to provide quantitative analysis [5]. Bibliometrics based identification of active authors and institutions has many benefits, i.e. helping students and researchers to identify active and relevant institutes for their area of interest, and enabling employers to recruit the most qualified potential researchers [3].

In various fields of science, e.g., in medicine, physics and social sciences, it is common to identify the highly cited papers [6], [7], [8]. Bibliometrics and citation analysis studies have also been conducted in software engineering, computer science and other disciplines, e.g., [4], [2], [3], [9], [10], [11], [12], [13], [14]. The highly cited papers usually provide insights into new avenues of research, a significant summary of

the state-of-the-art in a research area and a measure of scientific activity, in general [1], [2]. One of the key outlets for Agile research, “*Agile Software Development Conference (XP)*”, has not been evaluated under the lens of citation analysis alone or as a sub-field of its own (processes). XP Conference (“*International Conference on Extreme Programming (XP)*” - formerly “*Conference on Agile Software Development (AGILE)*”) was included in a bibliometrics study of Karanatsiou et al. [14] in the general domain of software engineering (where XP conference was the only process oriented conference in that study). The study of Chuang et al. [13] assessed agile software development, in general, for 221 published primary articles on the topic.

The purpose of this study is to provide an overview of the literature published in all XP conference proceedings. This study helps readers to understand the development and evolution of the XP conference from three main aspects: (i) the citation landscape and the most cited papers, (ii) the most active authors, institutions and countries, in terms of number of publications, and (iii) the identification of emerging research topics in XP conference publications and use of indexed keywords.

This paper is organized as follows: First, we discuss the research method and the data extraction technique. Second, we present the results of the analysis including findings on active individuals and institutes, highly cited papers and authors, and trends in the covered topics. Third, we discuss the threats to validity of the study. Finally, we summarize the findings and provide recommendations for future research.

## II. RESEARCH METHOD AND DATA EXTRACTION

The research data were collected from Scopus<sup>1</sup> database on September 2<sup>nd</sup>, 2018. Scopus is claimed to be the largest abstract and citation database of peer-reviewed literature. Scopus

<sup>1</sup><https://www.elsevier.com/solutions/scopus>

TABLE I  
SEARCH QUERIES FOR EXTRACTING PAPERS FROM SCOPUS

No.	Query String and its explanation	Papers
1	<b>CONF("XP") AND (LIMIT-TO(DOCTYPE,"cp"))</b> Select XP conference and conference papers only	758
2	<b>SRCTITLE("Lecture Notes in Business Information Processing" AND VOLUME(77) AND (LIMIT-TO(PUBYEAR,2011) AND (LIMIT-TO(DOCTYPE,"cp"))</b> Select "Lecture Notes in Business Information Processing" and only vol. 77 which includes conference papers for XP 2011	31

also provides citation data and allows to save the search results to a csv-file, for further analysis.

We started with the search string "1" (see Table I), to collect data related to all published XP conference papers. The search resulted in 758 papers. To our surprise, the search string "1" did not retrieve papers for the year 2011. We learned that the papers for the year 2011 do not include the information about the XP conference in the Scopus database. Thus, to collect those missing papers <sup>2</sup>, we complemented the findings with the search string "2", resulting in 31 papers.

The complete search gave us 789 papers (758+31), covering the years of 2002-2018 (published by September 2<sup>nd</sup>, 2018). The data, including e.g., names of the authors, title, publication year, source title, number of citations, link and abstract, were stored as a csv-file. We were also able to extract data from Scopus, directly, for the analysis of the affiliations and countries related to the authors (analysis of the search results in Scopus) as well as the top 20 cited papers (overview of the citations in Scopus). We used both MS Excel and R/RStudio for analyzing statistics and trends from the data.

### III. RESULTS

In 2001, the first "XP Universe" hosted tutorials, lectures, panel discussions, posters, workshops, and other less traditional discussions. A year later, the 2<sup>nd</sup> "XP Universe" and 1<sup>st</sup> "Agile Universe" were brought together to attract software experts, educators, and developers<sup>3</sup>, in general. In 2003 and 2004, the two conferences, "Extreme Programming and Agile Methods - XP/Agile Universe" and "Extreme Programming and Agile Processes in Software Engineering" were organized separately, but reported together in a Springer database. In 2005, the conferences were merged and formed a single venue: "Extreme Programming and Agile Processes in Software Engineering". Since 2007, the conference has been called as "Agile Processes in Software Engineering and Extreme Programming".

The Scopus database search yielded 789 papers in the proceedings of XP conference published between 2002 and 2018, see Fig. 1. The high number of papers for 2004 (n=96) is explained by the fact that the two aforementioned conferences are recorded together. The first five years of the XP conference

TABLE II  
TOP 20 COUNTRIES WITH MOST PAPERS (2002-2018)

Country	Papers	Country	Papers
United States	116	Brazil	25
United Kingdom	110	Austria	20
Italy	81	Netherlands	20
Finland	66	Spain	20
Sweden	61	Denmark	15
Norway	58	Australia	14
Canada	57	Israel	13
Germany	50	Poland	13
Ireland	37	Switzerland	12
New Zealand	39	Belgium	8

TABLE III  
AFFILIATIONS WITH MINIMUM 15 PAPERS

Affiliation	Papers
University of Calgary	39
Free University of Bozen-Bolzano	29
Universita degli Studi di Cagliari	28
SINTEF, Norwegian Inst. of Tech.	22
Victoria University of Wellington	20
Norges Teknisk-Naturv. Universitet	19
Chalmers University of Technology	17
SINTEF Digital	16
Open University, Walton Hall	16

cover about 38% and the first 10 years cover nearly 70% of all those papers. In XP conference, the average number of papers per year is 46.4 with a standard deviation of 20.3, using STDEV.P. The lowest number of papers are from year 2012 (n=15). The low number of paper may be an indication of rigorous selection process. Alternatively, some of the volumes include only research papers and short papers, whereas, some include e.g., abstracts of the posters or the position papers of the PhD symposium. Such variations are quite normal in various publication forums. The topmost values in Fig. 1 are the values from Scopus and the values at the bottom represent the number of accepted full papers retrieved from the prefaces of the conference books. Two of the conferences (XP2014 & XP2012) did not report the number of submitted full papers, but for those that had the information available, the acceptance rate was between 20% (XP2011) and 49% (XP2003), arithmetic mean of the rates being 32%.

The analysis from the Scopus data shows that majority of the XP conference papers originated from the United States (116), United Kingdom (110), Italy (81) and Nordic countries (Finland (n=66), Sweden (n=61) and Norway (n=58)), see Table II. It seems that these countries have a strong culture of agile in software development which is actively reported in XP conference. Table III shows the most frequent contributing institutions in XP conference, in which the top three are University of Calgary, Canada, Free University of Bozen-Bolzano, Italy and Università degli Studi di Cagliari, Italy. It is notable, however, that the number of countries and affiliations is related to the number of related authors for each paper. The study of Chuang et al. [13] did not report the total number of papers per country, but reported the top publishing institutions to be from the United States, Norway and United Kingdom.

<sup>2</sup><https://link.springer.com/book/10.1007/978-3-642-20677-1#toc>

<sup>3</sup><http://www.xpuniverse.com/>

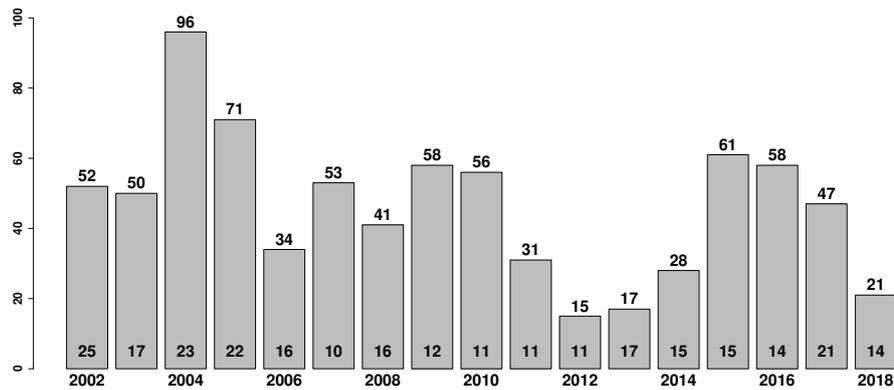


Fig. 1. Distribution of the Publications in Scopus (2002-2018) vs. total number of full papers in the conferences (value at the bottom)

### A. Authorship Trends

The results show that 1260 unique authors contributed to the 789 papers in XP conferences until 2018. The minimum number of authors for a XP paper was one whereas maximum was nine. Majority of the XP papers in 2018 (almost 35%) have four authors. In general, about 30% of all papers have two authors, 25% have one author, and 9% of the papers have five or more authors, see Table IV. The number of authors having contributed to three or more XP papers is rather small, as most authors have contributed to just one or two papers. About 75% of the authors (944) have an authorship to just one paper and about 88% of the authors (1108) have an authorship to only one or two papers, as a single or as a co-author. Chuang et al. [13] also reported a finding of a core intellectual pool contributing to the agile research realm.

During the first three years (2002-2004) of the conference, most papers were published by a single author. For the years 2005-2009, most papers were published by two authors, and for the years 2010-2012 and 2013-2014 by three and four authors, respectively. We consider the different number of authors for the papers as an indication of increased, high (international) collaboration among the contributors. In the 1970's, the average number of authors per paper in software engineering was around 1.5, while after 2010, the number of authors has typically been three [15]. The average number (i.e., arithmetic mean) of authors for the papers in XP conference is 2.6.

Asknes [16] studied a body of Norwegian articles (nearly 50000 articles having at least one Norwegian address). He concluded that at an aggregated, general level the “*highly cited papers typically involve more collaborative research than what is the normal or average*” [16]. In our study, the correlation between the number of authors and citations for a paper, for all papers, is weak ( $r = 0.13$ ,  $df = 787$ ,  $p = 0.0002$ ). However, for the set of top 20 cited papers (see Table VI), the correlation between the number of authors and citations for a paper is 0.59 ( $r = 0.59$ ,  $df = 18$ ,  $p = 0.0064$ ). Thus, the correlation coefficient suggests a strong positive correlation between the number of authors and citations for those top 20 cited papers.

Table V includes the 16 most active authors in the XP conference who have minimum number of 10 papers each. Maurer F. has been the most active author compared to the other top contributors of the XP conference. There are four authors that have their most cited papers published in 2010's (the publication year for the most cited paper in parenthesis), namely Abrahamsson P. (2015), Wang X. (2015), Concas G. (2012) and Bosch J. (2012); the rest of those most cited papers have been available for ten years or more. Interestingly, in a study “*Institutions, scholars and contributions on agile software development (2001-2012)*” by Chuang et al. [13], the list of the 18 most active authors included four of the 20 most active authors in this study, namely Abrahamsson P., Dingsøy T., Moe, N.B. and Sharp H. However, the list of the most active authors in that study [13] included also Boehm, B., Robinson H., Williams L., Dingsøy T., Moe, N.B. and Sharp H. who were among the authors of the top 20 most cited papers in this study.

### B. Citation Landscape & Most Cited Papers of XP Conference

A high citation count of a scientific work is an indication of the influential work and impact of a given paper [16], [17]. Our analysis shows that 62% ( $n=488$ ) of XP papers have been cited at least once, leaving about 38% ( $n=301$ ) as uncited papers, see Fig. 2. This is an indication of higher visibility of XP conference papers. When focusing on the first ten years of XP conference, i.e., the papers prior to 2012, nearly 65% of those papers (352/542) have been cited at least once. The findings are in line with prior studies [4], [18] in which about 43% of the papers were uncited (large body of software engineering publications). Similarly, about 42% of the papers of “*International Symposium on Empirical Software Engineering and Measurement*” [3] were uncited.

Garfield [1] argues about the citation count being the measure of *importance* or *impact* of a scientific work. He claims that citation count is rather a measure of *utility*, i.e., usefulness of the work for a large number of people or experiments [1]. Furthermore, a citation count can also be a measure of *scientific activity* and not necessarily related to the significance of the scientific work [1]. As in reality, only a

TABLE IV  
PROPORTION OF THE NUMBER OF THE AUTHORS PER YEAR

Year	Number of Authors								
	1	2	3	4	5	6	7	8	9
2002	<b>46.2%</b>	30.8%	9.6%	1.9%	5.8%	1.9%	1.9%	0.0%	1.9%
2003	<b>44.0%</b>	30.0%	12.0%	6.0%	2.0%	4.0%	2.0%	0.0%	0.0%
2004	<b>41.7%</b>	32.3%	13.5%	7.3%	1.0%	3.1%	1.0%	0.0%	0.0%
2005	26.8%	<b>36.6%</b>	21.1%	9.9%	2.8%	0.0%	1.4%	1.4%	0.0%
2006	14.7%	<b>32.4%</b>	17.6%	26.5%	8.8%	0.0%	0.0%	0.0%	0.0%
2007	22.6%	<b>37.7%</b>	15.1%	17.0%	7.5%	0.0%	0.0%	0.0%	0.0%
2008	9.8%	<b>39.0%</b>	29.3%	9.8%	4.9%	0.0%	4.9%	0.0%	2.4%
2009	24.1%	<b>34.5%</b>	19.0%	13.8%	3.4%	1.7%	3.4%	0.0%	0.0%
2010	25.0%	19.6%	<b>33.9%</b>	17.9%	1.8%	1.8%	0.0%	0.0%	0.0%
2011	16.1%	19.4%	<b>41.9%</b>	19.4%	3.2%	0.0%	0.0%	0.0%	0.0%
2012	6.7%	33.3%	<b>53.3%</b>	0.0%	6.7%	0.0%	0.0%	0.0%	0.0%
2013	5.9%	23.5%	17.6%	<b>47.1%</b>	5.9%	0.0%	0.0%	0.0%	0.0%
2014	14.3%	21.4%	17.9%	<b>25.0%</b>	10.7%	10.7%	0.0%	0.0%	0.0%
2015	13.1%	<b>27.9%</b>	23.0%	21.3%	6.6%	3.3%	1.6%	3.3%	0.0%
2016	25.9%	<b>27.6%</b>	25.9%	8.6%	5.2%	6.9%	0.0%	0.0%	0.0%
2017	10.6%	21.3%	<b>29.8%</b>	23.4%	8.5%	4.3%	2.1%	0.0%	0.0%
2018	9.5%	23.8%	23.8%	<b>33.3%</b>	9.5%	0.0%	0.0%	0.0%	0.0%
Total	24.7%	29.8%	21.8%	14.6%	4.8%	2.4%	1.3%	0.4%	0.3%

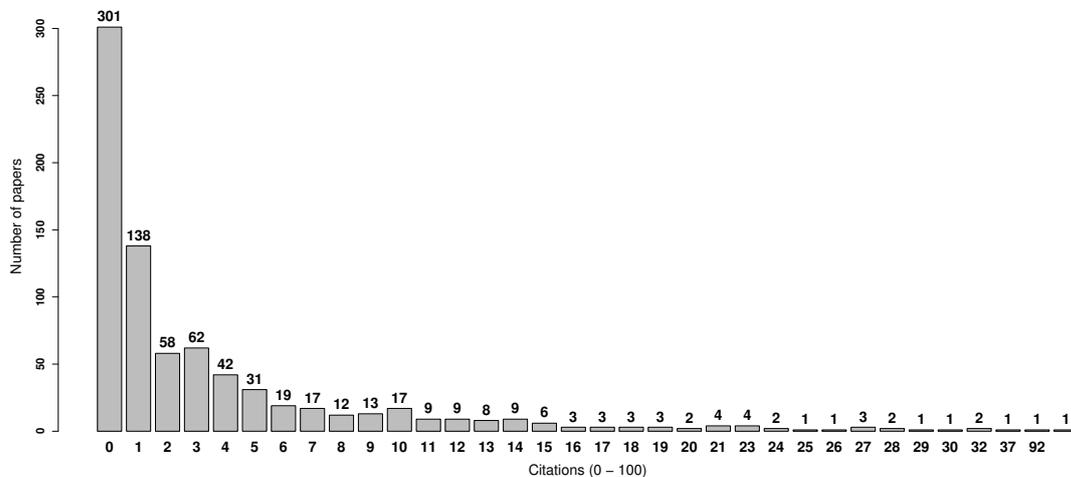


Fig. 2. Distribution of Citations (0-100) for the papers

rather small portion of the XP conference papers retrieved from Scopus are full research papers, the high number of uncited papers is not a surprise. Thus, it can be claimed that the samples from indexed databases may not be as representative as expected for citation analysis without rigorous filtering. However, such sample papers may well be valid for analysing author activity as well as research trends and topics.

The Table VI shows the top 20 most cited XP conference papers (each paper having minimum 23 citations). The total number of citations for the top 20 papers covers almost 25% of all citations (680/2920) which are mainly from earlier years of XP conference (2002-2009). However, one paper is published in 2015 and five papers among the top 20 papers are published in 2002. Table VI, shows that 92% of citations (624/680) are from papers not written by the authors (of the cited paper) themselves. Typically, a paper is cited the first time during the year of its publication or during the following year. However, the two top cited papers, “*Empirical findings*

*in agile methods*” by Lindvall et al. (2002) and “*Towards a framework for integrating agile development and user-centred design*” by Chamberlain et al. (2006), have been published over ten years ago, and have received the most citations since 2015. Chamberlain et al. (2006) had only a few citations right after its publication. After 2010 until 2015 the paper has received attention from both industry and academics in various fields of science, e.g., Computer Science, Mathematics, Decision science, Business, Management and Accounting, Social sciences or Psychology. In 2017, Chamberlain et al. (2006) received the most citations among the top 20 cited papers, and was the second most cited in 2018 (after Lindvall et al. 2002), at the time of the study.

### C. Highest Cited Papers Per Year

Many countries and evaluating bodies (for funding, promotions or appointments) are using figures like publication record or citation count in decision-making [3]. Such evaluations have two sides; firstly, it is fair to see the influential and trendy work

TABLE V  
MOST ACTIVE AUTHORS WITH MINIMUM 10 PAPERS

Author	#	Years (Papers)	Citations			1 <sup>st</sup> or 2 <sup>nd</sup> author <sup>c</sup>	
			Total	Avg.	Max <sup>a</sup>		
Maurer F.	29	2011 (2), 2010 (4), 2009 (5), 2008 (5), 2007 (6), 2006 (2), 2005 (1), 2004 (1), 2002 (3)	178	6.14	27 (2007)	6.10	17 (29)
Abrahamsson P.	18	2017 (3), 2016 (2), 2015 (2), 2014 (1), 2013 (1), 2009 (4), 2008 (2), 2007 (1), 2005 (1), 2004 (1)	85	4.72	21 (2015)	2.91	8 (18)
Marchesi M.	17	2018 (1), 2016 (2), 2015 (2), 2014 (1), 2013 (1), 2012 (1), 2011 (2), 2008 (1), 2007 (3), 2006 (1), 2004 (1), 2003 (1)	113	6.65	29 (2004)	3.87	5 (17)
Fraser S.	16	2015 (2), 2010 (1), 2009 (1), 2008 (1), 2007 (1), 2006 (2), 2005 (2), 2004 (2), 2003 (3), 2002 (1)	26	1.63	8 (2003)	0.89	16 (16)
Wang X.	14	2017 (3), 2016 (1), 2015 (2), 2014 (2), 2013 (1), 2010 (1), 2009 (2), 2008 (1), 2006 (1)	56	4.00	21 (2015)	1.92	7 (14)
Noble J.	13	2015 (1), 2014 (1), 2013 (1), 2012 (1), 2011 (2), 2010 (3), 2009 (1), 2008 (1), 2007 (1), 2004 (1)	105	8.08	28 (2007)	3.60	12 (13)
Sharp H.	13	2018 (1), 2017 (1), 2015 (1), 2014 (1), 2012 (1), 2011 (1), 2010 (2), 2008 (1), 2006 (2), 2005 (1), 2004 (1)	215	16.54	92 (2006)	7.36	10 (13)
Concas G.	12	2014 (3), 2013 (1), 2012 (1), 2011 (2), 2008 (1), 2007 (2), 2006 (1), 2005 (1)	69	5.75	14 (2012)	2.36	9 (12)
Dingsøy T.	12	2018 (3), 2017 (1), 2016 (1), 2015 (2), 2013 (1), 2011 (1), 2009 (2), 2008 (1)	71	5.92	32 (2008)	2.43	7 (12)
Holcombe M.	12	2008 (1), 2005 (8), 2004 (1), 2003(2)	19	1.58	7 (2005)	0.65	8 (12)
Succi G.	12	2011 (2), 2009 (3), 2008 (1), 2007 (2), 2005 (2), 2004 (1), 2003 (1)	52	4.33	18 (2008)	1.78	4 (12)
Bosch J.	11	2018 (1), 2017 (3), 2016 (1), 2015 (3), 2014 (2), 2012 (1)	36	3.27	15 (2012)	1.23	6 (11)
Hussman D.	11	2008 (1), 2007 (2), 2006 (1), 2005 (2), 2004 (5)	4	0.36	1 (2005)	0.14	6 (11)
Martin A.	11	2017 (1), 2008 (1), 2007 (1), 2006 (1), 2005 (3), 2004 (3), 2003 (1)	28	2.55	12 (2005)	0.96	10 (11)
Moe N.B.	10	2017 (2), 2016 (1), 2015 (1), 2013 (1), 2012 (1), 2011 (1), 2009 (2), 2008 (1)	71	7.1	32 (2008)	2.43	10 (10)
Mugridge R.	10	2005 (5), 2004 (3), 2003 (2)	16	1.60	5 (2003)	0.55	8 (10)

<sup>a</sup> Maximum number of citations for a single paper & publication year of that paper

<sup>b</sup> Percentage of the total number of citations (2920 for all publications)

<sup>c</sup> Number of times as first or second author in the publications

# Total number of publications

of specific investigator, and secondly, the appropriateness of such trends/counts can be questioned on scientific grounds. Rapid growth of citations for a paper may be a sign of a popular topic, or active author(s) building on their existing research, or both. Eight of the year-wise most cited papers are the same as reported in Table VI. Those papers have been available for the public for a long period of time, from years 2002 (5), 2004 (3), 2005 (1), 2006 (4), 2007 (3), 2008 (2), 2009 (1) and 2015 (1). The average number of citations for

top cited paper per year in Table VII is 26.6, which is less than the average from top 20 most cited papers, 34 in Table VI.

To compare the general interest on the published papers, we normalized the number of citations for years, see column C-Norm in Table VII. The values for normalized citations varied between 0.53–7.67. The highest number of normalized citations, 7.67, are for the paper “*What do practitioners vary in using scrum*” by Diebold et al. (2015) which received 23 citations in three years (ranked #8 in Table VII considering

TABLE VI  
TOP 20 CITED PAPERS (2002-2018, SORTED BY THE COLUMN "ALL")

#	Author(s) & Title (Year)	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	All <sup>a</sup>	NSC
1	Lindvall, M., Basili, V., Boehm, B., Costa, P., Dangle, K., Shull, F., Tesoriero, R., Williams, L., Zelkowitz, M.: Empirical findings in agile methods (2002)	1	2	4	8	8	6	6	5	6	8	6	6	4	6	9	9	6	100	96
2	Chamberlain, S., Sharp, H., Maiden, N.: Towards a framework for integrating agile development and user-centred design (2006)						1	2	2	5	8	9	7	17	12	10	14	5	92	89
3	Baheti, P., Gehring, E., Stotts, D.: Exploring the efficacy of distributed pair programming (2002)	1	4	2	2	2	3	3	5	7	2	1	2	1	3	1	0	0	37	33
4	Moë, N.B., Dingsy, T.: Scrum and team effectiveness: Theory and practice (2008)								1	3	4	3	5	2	6	4	4	0	32	28
5	Robinson, H., Sharp, H.: The characteristics of XP teams (2004)				4	3	1	3	3	2	1	6	3	3	1	1	1	1	32	28
6	Turner, R., Jain, A.: Agile meets CMMI: Culture clash or common cause? (2002)			1	2	1	2	3	2	1	4	4	1	4	1	4	2	2	30	30
7	Mannaro, K., Melis, M., Marchesi, M.: Empirical analysis on the satisfaction of IT employees comparing XP practices with other software development methodologies (2004)	2				1	1	1	4	4	2	3	2	2	2	3	4	2	29	27
8	Ferreira, J., Noble, J., Biddle, R.: Up-front interaction design in agile development (2007)						2	3	2	2	5	2	4	1	2	3	2	2	28	24
9	Stetos, P., Stamelos, I., Angelis, L., Deligiannis, I.: Investigating the impact of personality types on communication and collaboration-viability in pair programming - An empirical study (2006)					1	1	2	2	5	2	3	2	3	3	1	2	1	28	23
10	Abbas, N., Gravel, A.M., Willis, G.B.: Historical roots of agile methods: Where did "Agile thinking" come from? (2008)									2	3	4	4	1	5	3	2	3	27	27
11	Tessier, B., Maurer, F.: Job satisfaction and motivation in a large agile team (2007)						1	2		3	5	2	4	3	2	3	2	2	27	26
12	Stotts, D., Lindsey, M., Antley, A.: An informal formal method for systematic junit test case generation (2002)	1	3	5	7			4	2	1	1	1		1	1		1	0	27	27
13	Bryant, S., Romero, P., Duboulay, B.: The collaborative nature of pair programming (2006)						1	1	2	4	3	5	1	4	3	2	0	0	26	25
14	Hussain, Z., Milchrahm, H., Shahzad, S., Siany, W., Tscheligi, M., Wolkerstorfer, P.: Integration of extreme programming and user-centered design: Lessons learned (2009)							2	3	2	4	2	4	1	5	2	0	0	25	21
15	Haikara, J.: Usability in agile software development: Extending the interaction design process with personas approach (2007)									3	2	3	2	3	1	4	5	1	24	24
16	Middleton, P., Flaxel, A., Cookson, A.: Lean software management case study: Timberline Inc. (2005)									1	4	3	5	3	4	1	2	1	24	23
17	Diebold, P., Ostberg, J.P., Wagner, S., Zender, U.: What do practitioners vary in using scrum? (2015)													4	5	10	4	23	8	
18	Melmik, G., Maurer, F.: Comparative analysis of job satisfaction in agile and non-agile software development teams (2006)										3	3	3	5	4	1	3	3	23	22
19	Koch, S.: Agile principles and open source software development: A theoretical and empirical discussion (2004)						1	1	5	5	4	3				1	2	1	23	21
20	Melmik, G., Maurer, F.: Perceptions of agile practices: A student survey (2002)		1	1	4	1	1	1	4	2	5	1	2				1	0	23	22

<sup>a</sup>All = All Citations (including self-citations), NSC = All Citations excluding self-citations

purely citations). Similarly, the paper “*Empirical findings in agile methods*” by Lindvall et al. (2002) has been available for twelve years and has 92 citations (similarly, ranked as #2 in Table VII). The paper also ranked the highest for the number of citations (100, see Table VI) and has the fourth highest normalized citation count (6.25).

Garousi and Fernandes [18] claim that newer papers will first get to be known in the community. According to Raulamo-Jurvanen et al [3] the longer the paper has been available the better are the chances to be cited. However, according to our results, recent papers have received more attention in terms of citations. One reason can be that the software engineering community has grown over the years and recent topical papers may have a slight advantage when it comes to the number of citations per year.

We were curious to see whether the length of the title had impact on the number of citations for a paper. Letchford et al. [19] had studied the relationship between the lengths of paper titles and citations (across various journals) and concluded that a short title for a paper is an advantage for receiving citations. However, they also stated that the evidence is not as strong when adjusted for the journal where the paper is published. For the XP papers, the correlation between the length of the title, either in words or characters, and the number of citations is weak ( $r = 0.03$ ,  $df = 787$ ,  $p = 0.415$  and  $r = 0.04$ ,  $df = 787$ ,  $p = 0.235$ , respectively). The top 5 cited papers have rather short titles (length varying from 31 to 77 in characters and from 5 to 10 in words). The median length of all titles, in characters and words is 62 and 8, respectively.

#### D. Topical Issues

With topic modeling, we intend to analyze the abstract topics in the documents. We removed 66 documents from the original pool of 789 documents, as not including the abstract in Scopus. Thus, the set of documents for trend analysis included 723 documents. We combined the titles and the abstracts of the documents, converted the text to lowercase and removed all (english) stopwords in R.

For the trend analysis we utilized topic modeling and Latent Dirichlet Allocation (LDA) as described by Griffiths and Steyvers [12] with R scripts based on Ponweiser [20]. Our approach was identical to the process used by Raulamo-Jurvanen et al. [3] and Garousi and Mäntylä [4]. We created a document term matrix from the corpus (using R “text2vec”<sup>4</sup> package), excluding words having less than two characters or appearing in less than three documents. We generated a LDA model (using R “topicmodels”<sup>5</sup> package) by running the topic models from 2 to 100 by one, yielding 35 as the optimal number of topics.

In the analysis of the trend slopes (by publication year) the topics gaining interest among the authors are the “hot topics” and the topics declining interest are the “cold topics”. The five hottest and coldest topics, interpreted by the topic-specific

words (and related titles), and 10 significant terms for each of those, as shown in Table VIII(a) and Table VIII(b), respectively. The topics gaining the most interest are “Coordination” and “Technical Debt”, which include issues like largescale coordination and interteam objectives as well as metric and automation. Cold topics such as “Education”, “Methods and Practices” (including pair programming) and “Testing”, have been of less inspiration for the submissions during the recent years of XP conference.

In 2012, Dingsøy et al. [21] studied agile software development and outlined key research themes at the time, namely *Case Study Methodology, Traditional Software Engineering, CMM, Project Management, Software estimation, Pair Development, Distributed Cognition, Agile methods, User-centered design, Agile methodologies* and *Patterns*. Some of those themes seem still topical, e.g., *software estimation* as “Technical Debt” and some not, like *Pair Development* or *Agile Methods* as “Methods and Practices” (see Table VIII). In fact, Dingsøy et al. [21] report that in Agile2011 they had specifically asked people (mainly academics) what are the topics that should be researched less or further. Pair programming in educational settings and reuse of code were considered as topics not requiring further research while topics like agile across projects and across organizations and distributed agile were considered to be important. “*We concur that these are exciting research areas that can further our understanding of the effectiveness of agile methods and practices, particularly in different project/organizational contexts*” [21]. Such trend is also visible in our study, as “Education” and “Methods and Practices” (including pair programming) were found to be cold topics and topics like “Coordination” and “Teamwork” were among the hot topics.

Perhaps researchers should ask research topic related questions more frequently, not only among academics but also among the practitioners in the field, to support the needs or interests in the industry, too.

#### E. Indexed Keywords

To study the published topics from another perspective, we collected the indexed keywords from Scopus. It is notable that we used the indexed keywords (not the author keywords), as the indexed keywords outnumber the author keywords, providing more details. Additionally, there are papers that are not only missing abstracts (see Chapter III-D) but also keywords (see Scopus e.g., a conference paper “*Agile acceptance testing*” by Pettichord and Marick from 2002). There were 720 papers with indexed keywords. The minimum number of indexed keywords for a paper was 3, the maximum was as high as 25 (for one paper) and arithmetic mean 9.4. We checked the correlation between the number of indexed keywords and the number of citations for a paper, but that correlation is weak ( $r = 0.028$ ,  $df = 718$ ,  $p = 0.459$ ).

We paired the keywords for each paper (e.g., a paper having four keywords would eventually yield 6 unique keyword pairs) and converted the keywords to lower case. The pairing resulted in 32131 keyword pairs which we then stored in a CSV-file.

<sup>4</sup><https://cran.r-project.org/web/packages/text2vec/index.html>

<sup>5</sup><https://cran.r-project.org/web/packages/topicmodels/index.html>

TABLE VII  
TOP CITED PAPERS PER YEAR (2002-2018)

Year	Author(s) & Title	Cites	C-Norm	Rank
2002	Lindvall,M., Basili,V., Boehm,B., Costa,P., Dangle,K., Shull,F., Tesoriero,R., Williams,L., Zelkowitz,M.: Empirical findings in agile methods	100	6.25	4
2003	Lowell C.,Stell-Smith J.:Successful automation of GUI driven acceptance testing	8	0.53	17
2004	RobinsonH.,SharpH.: The characteristics of XP teams	32	2.29	12
2005	Middleton,P., Flaxel,A., Cookson,A.: Lean software management case study: Timberline Inc.	24	1.85	14
2006	Chamberlain,S., Sharp,H., Maiden,N.: Towards a framework for integrating agile development and user-centred design	92	7.67	1
2007	Ferreira,J., Noble,J., Biddle,R.: Up-front interaction design in agile development	28	2.55	10
2008	MoeN.B.,DingsyrT.: Scrum and team effectiveness: Theory and practice	32	3.20	7
2009	Hussain,Z., Milchrahm,H., Shahzad,S., Slany,W., Tscheligi,M., Wolkerstorfer,P.: Integration of extreme programming and user-centered design: Lessons learned	25	2.78	9
2010	FerreiraJ.,SharpH.,RobinsonH.: Values and assumptions shaping Agile development and User Experience design in practice	14	1.75	15
2011	DorairajS.,NobleJ.,MalikP.: Effective communication in distributed agile software development teams	15	2.14	13
2012	StaronM.,MedingW.,PalmK.: Release readiness indicator for mature agile and lean software development projects	21	3.50	5
2013	HeikkiläV.T.,PaasivaaraM.,LasseniusC.,EngblomC.: Continuous release planning in a large-scale scrum development organization at ericsson	12	2.40	11
2014	LiskinO.,PhamR.,KieslingS.,SchneiderK.: Why we need a granularity concept for user stories	12	3.00	8
2015	Diebold,P.,Ostberg,J.-P.,Wagner,S.,Zendler,U.: What do practitioners vary in using scrum?	23	7.67	1
2016	OrtuM.,DestefanisG.,CounsellS.,SwiftS.,TonelliR.,MarchesiM.: Arsonists or firefighters? Effectiveness in agile software development	7	3.50	5
2017	TaibiD.,LenarduzziV.,JanesA.,LiukkunenK.,AhmadM.O.: Comparing requirements decomposition within the Scrum, Scrum with Kanban, XP, and Banana development processes	7	7.00	3
2018	OyetoanT.D.,MiloshekaB.,GriniM.,SoaresCruzesD.: Myths and facts about static application security testing tools: An action research at telenor digital	1	1.00	16

<sup>a</sup> C-Norm = Citations divided by the number of years a paper has been available

TABLE VIII  
HOT AND COLD TOPICS, TERMS & NUMBER OF PAPERS FOR EACH TOPIC

(a) Hot Topics				
Coordination	Technical Debt	Teamwork	Startups	Agile Practices
24	21	23	18	30
largescale	technical	meeting	startup	scrum
coordinate	debt	retrospective	devops	kanban
mechanism	metric	reflection	prototype	board
tailor	evolution	standup	stage	barriers
interteam	td	commitment	speed	wip
userstory	production	workshop	sprints	selforganizing
standard	automatic	education	monitoring	multitasking
story	stakeholders	scalability	pressure	automotive
objectives	monitored	guideline	theoretical	optimization
human	influencing	enhance	attempts	transformations
(b) Cold Topics				
Process Simulation	Education	Coaching & Experimenting	Testing	Methods and Practices
52	28	17	21	31
xp	student	coach	acceptance	pair
simulation	teach	languages	executable	programmer
integrate	university	transition	version	experiment
budget	education	mock	regulations	skill
units	curriculum	panel	workshop	tester
leadership	skill	standard	testdriven	switching
waterfall	classroom	tutorial	packages	assist
events	testable	certified	technical	standard
tester	selforganizing	exercises	classify	structures
userinterface	comprehensive	shares	methodological	expectations

We used the Cytoscape<sup>6</sup>, an open source software platform, for visualizing the network of the paired keywords (after removing duplicates), see Fig. 3. The lighter the color in the figure, the more the keyword had connections. The keyword “*software engineering*” was, unsurprisingly, the most used keyword, see Fig. 3. The nine other most used keywords were “*software design*”, “*agile software development*”, “*agile methods*”, “*computer programming*”, “*project management*”, “*computer software*”, “*agile development*”, “*extreme programming*”, “*agile*” and “*software testing*”. The keywords are rather generic, but still quite nicely represent the key research themes identified by Dingsøyr et al. [21]. However, a more detailed analysis of the keywords, to view the overall importance and reveal the topicality of the keywords, would be required to see the trends in the area of XP.

#### IV. THREATS TO VALIDITY

In this section, we discuss four perspectives of validity threats [17] and the steps that we have taken to mitigate those threats.

Internal validity reflects the extent to which a causal conclusion based on a study is warranted [17]. The approach used for the selection and extraction of XP conference paper from selected are discussed in Section II. In order to ensure repeatability and reproducibility of our study, the search terms have been defined carefully and reported in the research method Section II. Additionally, the raw data and the scripts used are provided to ensure transparency and replicability of our analysis. The material can be accessed via this link: <https://bit.ly/2LiqQ3S>.

Construct validity is concerned with issues that to what extent the object of study truly represents theory behind the study [17]. As a limitation w.r.t. construct validity, we assumed that all the papers were published in Scopus database properly. Scopus claims to be “*the largest abstract and citation database of peer-reviewed literature*”<sup>7</sup>. All the XP conference proceedings are indexed in Scopus and we fetched all the data from this database. However, 2011 papers are not properly indexed, so papers for the year of 2011 were fetched with a separate query and added to the research data manually.

Conclusion validity of a study deals with whether correct conclusions are reached through rigorous and repeatable treatments [17]. Throughout the paper, the discussions and conclusions are based on actual quantitative measures and statistics from the extracted data. The approach we used to identify and map the top papers assures that, the results of any replications of this study will not have major deviations from our results.

External validity is concerned with to what extent the results of this secondary study can be generalized [17]. The results of this study are not meant to be generalized to the whole SE field or outside SE. However, we believe that given the rigor of our approach that we used to identify top cited papers, emerging

hot topics, the results highlight the citation landscape of the top XP conference papers in SE area.

#### V. CONCLUSIONS AND FUTURE WORK

This is the first citation and topic analysis study on XP conference papers since 2002 until 2018. The paper identifies and classifies: the highly cited papers, topic trends, top individuals and institutes who have significantly published in XP conference.

The trend of the papers shows that XP conference has received interest from both the academic community and industry. The papers highlight that much of research is stirred by practices emerging in industry. Overall, 62% of the XP conference papers received at least one citation, which is a sign of good visibility relevance of the published papers. However, about 38% of the XP papers so far have received no citations at all. This raises concerns and questions such as: what are the reason(s) of large ratio of non-cited XP conference papers? Does this have anything to do with papers or venues quality? Or, is it about the topics of the papers, the indexed keywords, or the keywords provided by the author(s)? The data, which we make publicly available, can be used to conduct various analysis (i.e., characteristics of highly cited papers) on XP conference papers.

The analysis shows that XP community interest has been moving away from “Process Simulation”, “Education” and “Coaching & Experimenting” related topics to more practice and process oriented topics. According to the trend analysis, the hottest research topics, i.e., the topics gaining the most interest are “Coordination”, “Technical Debt”, “Teamwork”, “Startups” and “Agile Practices”. The identified trends are helpful for both researchers and practitioners to see topics that are more impact and align their future research activities.

The study found an active core intellectual pool of authors along with their highly cited work. The newbie researchers can start their journey from these papers and follow listed active researchers to stay up to date about latest trends in the Agile world. Additionally, the active publishing institutes in XP conference can be helpful for doctoral students to approach experts on the specific topic for further research and doctoral studies. We hope that this paper encourages further discussions in the software engineering community towards further analysis and formal characterization of the highly-cited software engineering papers in general and specifically in XP conference community. The important thing about citation count is that it is an “*objective measure of the utility or impact of the scientific work*” [1].

The following are among our future work directions:

- To replicate this analysis for other SE publication venues in order to conduct comparison between research venues and provide more depth to our analysis.
- To mine typical features for highly cited papers and to assess the extent to which papers inner quality, external features, reputation of the authors and journals, contribute to generation of highly cited papers in the future.

<sup>6</sup><https://cytoscape.org/>

<sup>7</sup><https://www.elsevier.com/solutions/scopus>

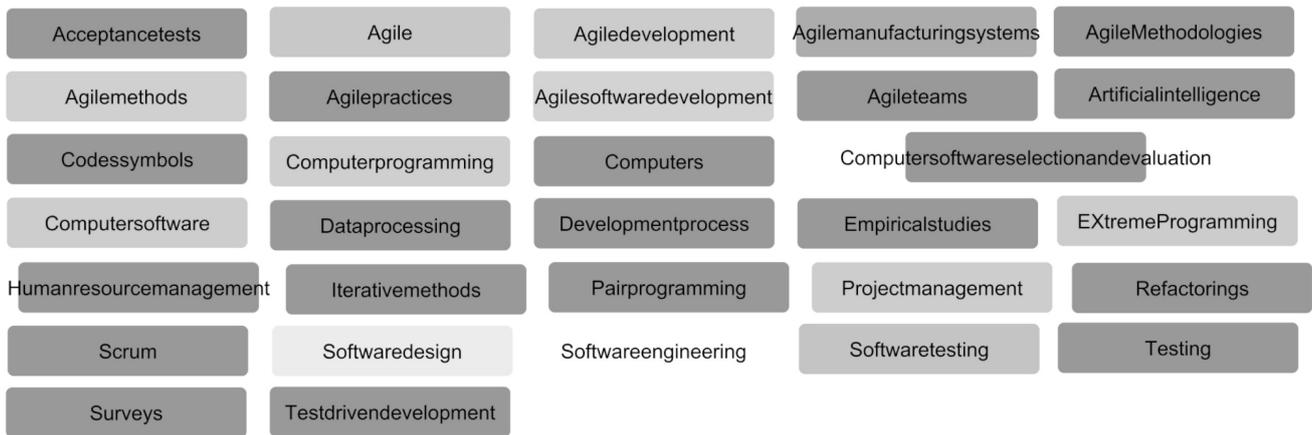


Fig. 3. Visualization of the most connected, paired indexed keywords (31)

- To study the indexed keywords within a publication venue, in more detail, e.g., by years, to see whether we could find trends from those, too.

#### REFERENCES

- [1] E. Garfield, "Is citation analysis a legitimate evaluation tool?" *Scientometrics*, vol. 1, no. 4, pp. 359–375, May 1979. doi: 10.1007/BF02019306. [Online]. Available: <https://doi.org/10.1007/BF02019306>
- [2] C. Wohlin, "An analysis of the most cited articles in software engineering journals - 2000," *Information and Software Technology*, vol. 49, no. 1, pp. 2–11, 2007. doi: <https://doi.org/10.1016/j.infsof.2006.08.004>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0950584906001133>
- [3] P. Raulamo-Jurvanen, M. V. Mäntylä, and V. Garousi, "Citation and topic analysis of the esem papers," in *2015 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, Oct. 2015. doi: 10.1109/ESEM.2015.7321193. ISSN 1949-3770 pp. 1–4.
- [4] V. Garousi and M. V. Mäntylä, "Citations, research topics and active countries in software engineering: A bibliometrics study," *Computer Science Review*, vol. 19, pp. 56–77, 2016. doi: <https://doi.org/10.1016/j.cosrev.2015.12.002>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1574013715300654>
- [5] OECD. (2013) *Oecd frascati manual*, sixth edition, annex 7, paras. 20–22, oxford dictionaries, 2013, website. [Online]. Available: <https://stats.oecd.org/glossary/detail.asp?ID=198>
- [6] T. A. Hamrick, R. D. Fricker, and G. G. Brown, "Assessing what distinguishes highly cited from less-cited papers published in interfaces," *INFORMS Journal on Applied Analytics*, vol. 40, no. 6, pp. 454–464, 2010. doi: 10.1287/inte.1100.0527. [Online]. Available: <https://pubsonline.informs.org/doi/abs/10.1287/inte.1100.0527>
- [7] R. Danell, "Can the quality of scientific work be predicted using information on the author's track record?" *Journal of the American Society for Information Science and Technology*, vol. 62, no. 1, pp. 50–60, 2011. doi: 10.1002/asi.21454. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.21454>
- [8] L. Bornmann, "How are excellent (highly cited) papers defined in bibliometrics? a quantitative analysis of the literature," *Research Evaluation*, vol. 23, no. 2, pp. 166–173, 03 2014. doi: 10.1093/reseval/rvu002. [Online]. Available: <https://dx.doi.org/10.1093/reseval/rvu002>
- [9] R. L. Glass, I. Vessey, , and V. Ramesh, "Research in software engineering: an analysis of the literature," *Information and Software Technology*, vol. 44, no. 8, pp. 491–506, 2002. doi: [https://doi.org/10.1016/S0950-5849\(02\)00049-6](https://doi.org/10.1016/S0950-5849(02)00049-6). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0950584902000496>
- [10] A. Hoonlor, B. K. Szymanski, and M. J. Zaki, "Trends in computer science research," *Commun. ACM*, vol. 56, no. 10, pp. 74–83, Oct. 2013. doi: [https://dx.doi.org/10.1016/S0950-5849\(02\)00049-6](https://dx.doi.org/10.1016/S0950-5849(02)00049-6)
- [11] R. V. Noorden, B. Maher, and R. Nuzzo, "The top 100 papers," *Nature*, vol. 514, no. 7524, pp. 550–553, 2014. doi: doi:10.1038/514550a
- [12] T. L. Griffiths and M. Steyvers, "Finding scientific topics," *Proceedings of the National Academy of Sciences*, vol. 101, no. suppl 1, pp. 5228–5235, 2004. doi: 10.1073/pnas.0307752101. [Online]. Available: [http://www.pnas.org/content/101/suppl\\_1/5228](http://www.pnas.org/content/101/suppl_1/5228)
- [13] S.-W. Chuang, T. Luor, and H.-P. Lu, "Assessment of institutions, scholars, and contributions on agile software development (20012012)," *Journal of Systems and Software*, vol. 93, pp. 84–101, 2014. doi: <https://dx.doi.org/10.1016/j.jss.2014.03.006>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0164121214000697>
- [14] D. Karanatsiou, Y. Li, E.-M. Arvanitou, N. Misirlis, and W. E. Wong, "A bibliometric assessment of software engineering scholars and institutions (20102017)," *Journal of Systems and Software*, vol. 147, pp. 246–261, 2019. doi: <https://doi.org/10.1016/j.jss.2018.10.029>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0164121218302334>
- [15] V. Garousi and J. M. Fernandes, "Highly-cited papers in software engineering: The top-100," *Information and Software Technology*, vol. 71, pp. 108–128, 2016. doi: <https://doi.org/10.1016/j.infsof.2015.11.003>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0950584915001871>
- [16] D. W. Aksnes, "Characteristics of highly cited papers," *Research Evaluation*, vol. 12, no. 3, pp. 159–170, 12 2003. doi: 10.3152/147154403781776645. [Online]. Available: <https://dx.doi.org/10.3152/147154403781776645>
- [17] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén, *Experimentation in Software Engineering: An Introduction*, ser. International Series in Software Engineering. Springer US, 2000, vol. 6.
- [18] V. Garousi and J. M. Fernandes, "Quantity versus impact of software engineering papers: a quantitative study," *Scientometrics*, vol. 112, no. 2, pp. 963–1006, Aug 2017. doi: 10.1007/s11192-017-2419-6. [Online]. Available: <https://dx.doi.org/10.1007/s11192-017-2419-6>
- [19] A. Letchford, H. S. Moat, and T. Preis, "The advantage of short paper titles," *Royal Society Open Science*, vol. 2, no. 8, pp. 1–6, 2015. doi: doi.org/10.1098/rsos.150266. [Online]. Available: <https://dx.doi.org/10.1098/rsos.150266>
- [20] M. Ponweiser, "Latent dirichlet allocation in r," Master's thesis, Institute for Statistics and Mathematics, University of Economics and Business, Vienna, Austria, 2012.
- [21] T. Dingsøyr, S. Nerur, V. Balijepally, and N. B. Moe, "A decade of agile methodologies: Towards explaining agile software development," *Journal of Systems and Software*, vol. 85, no. 6, pp. 1213–1221, 2012. doi: <https://doi.org/10.1016/j.jss.2012.02.033> Special Issue: Agile Development. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0164121212000532>

## Factors that contribute significantly to Scrum adoption

Ridewaan Hanslo  
Council for Scientific and  
Industrial Research  
Pretoria, South Africa  
Email: rhanslo@csir.co.za

Ernest Mnkandla  
University of South Africa, School  
of Computing, College of Science,  
Engineering and Technology  
Pretoria, South Africa  
Email: mnkane@unisa.ac.za

Anwar Vahed  
Data Intensive Research Initiative  
of South Africa  
Pretoria, South Africa  
Email: avahed@dirisa.ac.za

**Abstract**—Scrum is the most adopted Agile methodology. The research conducted on Scrum adoption is mainly qualitative and there is therefore a need for a quantitative study on Scrum adoption challenges. The primary objective of this paper is to present the findings of a study on the factors that have a significant relationship with Scrum adoption as perceived by Scrum practitioners working within South African organizations. Towards this objective, a narrative review to extract and synthesize the existing challenges was conducted. These synthesized challenges were used in the development of a conceptual framework for evaluating the challenges that have a correlation and linear relationship with Scrum adoption. Following this, a survey questionnaire was used to test and evaluate the factors forming part of the developed framework. The findings indicate that Relative Advantage, Complexity, and Sprint Management are factors that have a significant linear relationship with Scrum adoption. Our recommendation is that organizations consider these findings during their adoption phase of Scrum.

**Index Terms** — Adoption Challenges, Agile Methodologies, Diffusion of Innovation, Multiple Linear Regression, Narrative Review, Quantitative Research, Scrum.

### I. INTRODUCTION

SCRUM is regarded as one of the most under researched Agile methodologies [1], and the majority of research conducted in this field is qualitative in nature [2]. This paper focuses on bridging this literature gap between the body of qualitative knowledge on Scrum and the lack of sufficient quantitative literature on Scrum adoption within the South African (SA) context.

The author's previous paper on Scrum adoption challenges focused on developing a model that can be used to test and evaluate challenges to Scrum adoption [3]. To test and evaluate the Scrum adoption challenges a narrative review was conducted on the existing Agile and Scrum adoption challenges experienced globally and within SA. The synthesized challenges were used as the independent variables to the model. The first iteration of the Conceptual

Framework (CF) is known as the Scrum Adoption Challenges Detection Model (SACDM). The CF is a custom model adapted from the Diffusion of Innovation (DOI) theory and the study of the adoption of new technology by Sultan & Chan [12]. The model is divided into four constructs, namely, Individual Factors ( $X_1$ ), Team Factors ( $X_2$ ), Organizational Factors ( $X_3$ ), and Technology Factors ( $X_4$ ). The independent variables are the factors within the constructs  $X_1, X_2, X_3$  and  $X_4$ . The dependent variable is  $Y$  with  $Y=f(X_1, X_2, X_3, X_4)$ . When  $Y=1$ , the individual within an organization is an adopter of Scrum. When  $Y=0$ , the individual within the organization is a non-adopter of Scrum. The first iteration of the CF is similar to the second iteration except that the statistical analysis technique is modified from linear regression to logistic regression. For this reason, the first iteration is not depicted.

In the second iteration the statistical analysis technique used to evaluate the dependent variable changed from multiple logistic regression to multiple linear regression (MLR). The reason for this change was because of the need to test and evaluate whether there was a statistically significant linear relationship between the adoption challenges and Scrum adoption. Another reason was the small sample size which did not meet the requirement of a large sample size for logistic regression. Figure 1, displays the second iteration of the CF labelled as the Scrum Adoption Challenges Conceptual Framework (SACCF). Independent variables are depicted as factors within constructs  $X_1, X_2, X_3$  and  $X_4$ . The dependent variable is  $Y$  with  $Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \epsilon$ . The constants  $\beta_i$  are the standardized coefficients (beta), and  $\epsilon$  is the standard error. The hypothesized relationships between the independent variables and the dependent variable are shown by the symbols in parenthesis.

The third iteration is the final version of the CF. The statistical analysis technique for the second and third iteration is identical. The third iteration creates a new set of 14 validated factors from the second iteration's 19 factors. This iteration of the CF is discussed in Section V. A quantitative survey was conducted using an online survey questionnaire. A set of 207 valid responses to this survey

was used to perform Exploratory Factor Analysis (EFA), and Cronbach's alpha analysis, which confirmed the validity and reliability of the survey instrument used.

The results from the correlational and MLR statistics were used to identify factors which have a significant linear relationship with Scrum adoption.

This paper consists of the following sections: Section II describes the background of the topic. Section III presents the methodology, including the statistical analysis techniques used to analyze and validate the data collection instrument. Section IV displays the results, followed by a discussion of the research findings in Section V. Section VI concludes the paper.

## II. BACKGROUND

### a) Scrum Defined

Scrum is one of many Agile software development methodologies available. Scrum has seen exponential growth in the past decade [7]. As a framework, Scrum allows organizations to improve on their project delivery objectives [17]. The Scrum guide written by Ken Schwaber and Jeff Sutherland describes this framework as lightweight, simple to understand, but extremely difficult to master [8].

Scrum embodies iterative and incremental development, and the framework is comprised of six artifacts, five roles, and four predominant activities [8].

### b) Agile Challenges

The introduction of new methodologies typically poses challenges for individuals and organizations who make use of them [9]. The adoption of Agile methodologies creates additional challenges such as management style, software development process, and software developer resistance [2].

The Agile adoption challenges in the context of this paper is taken from the author's previous paper on the Scrum Adoption Challenges Detection Model (SACDM) [3]. The challenges were derived from Agile, Scrum, Software Development Methodology (SDM), and Information Systems (IS) literature. These challenges are encountered both within South Africa (SA) and globally (non-SA).

Due to Scrum research within SA being primarily qualitative in nature [10], other Agile methodology challenges were included in order to attain a more comprehensive model. Common challenges such as lack of experience, the Organizational Culture, and lack of communication have been identified during the narrative review.

### c) Theoretical Framework

Research by Chan and Thong [11], and Mohan and Ahlemann [9] explains that previous IT adoption studies focused on the technical aspects of the innovation. These studies made use of technology adoption models, such as Technology Adoption Model (TAM). However, with complex Agile methodologies such as Scrum where

collaboration between individuals within teams and organizations are important, a more inclusive model was required. The mixture of factors which affect adoption led to the selection the Diffusion of Innovation (DOI) theory as the theoretical lens for the Conceptual Framework (CF) [13].

The DOI theory is used in both organizational and individual adoption studies, with the DOI model composed of five characteristics of innovation. The five characteristics of innovation are Compatibility, Complexity, Observability, Relative Advantage, and Trialability [13].

In the authors' custom model, as shown in Figure 1, Compatibility, Complexity, and Relative Advantage are the three characteristics of innovation that have been retained. The reason for this decision was based on the consistency of the relationship between the three characteristics and adoption behavior as identified within innovation studies [14].

## III. METHODOLOGY

### a) Research Design

The research design consists of a narrative review and survey questionnaire. The narrative review is a literature review to assess a topics body of knowledge [15]. This review was conducted due to the lack of quantitative literature on Scrum adoption. The review extracted and synthesized the Scrum and Agile adoption challenges to form the factors of the Conceptual Framework (CF).

The quantitative survey design operationalized the narrative reviews factors as the independent variables and Scrum adoption as the dependent variable. The online survey was used as the scale to measure the opinions of the Scrum practitioners working within SA organizations [16].

The validity of the scale was tested using a pilot study, Exploratory Factor Analysis (EFA), Bartlett's test for Sphericity, and Kaiser-Meyer-Olkin (KMO). Bartlett's test for Sphericity, EFA, and KMO are discussed in the analysis subsection. For reliability the Cronbach's coefficient alpha was used to measure internal consistency of the scale [16].

### b) Analysis

EFA is a statistical method used to describe the variability of the observed variables in terms of the unobserved constructs [4]. The validation of the questionnaire items against the initial 19 factors in the SACCF required a first order and second order EFA to be conducted. In the first order EFA we considered the 78 survey questionnaire items to construct the newly validated 14 factors. These factors were subjected to a second order EFA in order to develop the four constructs. The validity analysis proceeded by generating the first order EFA scores. Once the first order EFA scores were summarized, the second order EFA followed.

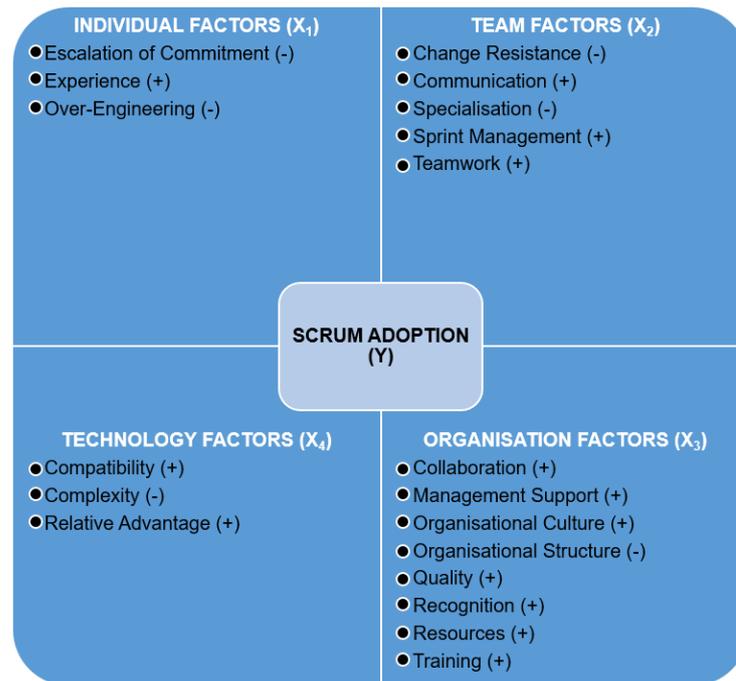


Figure 1: Scrum Adoption Challenges Conceptual Framework (SACCF).

To test the sampling adequacy, the KMO measure of sampling adequacy was used. The KMO value obtained was 0.88. The Bartlett’s test for Sphericity was conducted to determine if it was useful to conduct factor analysis. The Bartlett’s test for Sphericity significance level was 0.00. These test results indicate that it was, therefore, worthwhile to conduct the EFA on the dataset.

To determine the number of factors derived from the individual statements, Eigenvalues > 1 and the Scree plot were used. The constructs cumulative percentage was 75.8%.

The Principal Axis Factoring (PAF) extraction method with oblique rotation was used to seek a parsimonious representation for the common variance (correlation) between variables by latent factors. The oblique rotation implemented the Oblimin with Kaiser Normalization method because it was required to explore the correlations between the factors.

To summarize, of the 78 questionnaire items, 14 factors were retained for rotation due to their Eigenvalues being greater than or near one. The first 14 factors as a collective accounted for 75.8% of the total variance.

Because of the factor loading cut-off criteria of 0.40, 12 items were found to load on the first factor, and these were subsequently labelled "Organizational Behavior". Eight items loaded on the second factor, labelled "Sprint Management". Nine items loaded on the third factor, labelled "Relative Advantage". Four items loaded on the fourth, fifth, sixth, and the seventh factor respectively, labelled "Experience", "Training", "Specialization", and

"Recognition". Seven items loaded on the eighth factor, labelled "Customer Collaboration". Three items loaded on the ninth factor, labelled "Compatibility". Five items loaded on the tenth factor, labelled "Over-Engineering". Three items loaded on the eleventh and twelfth factor respectively, labelled "Escalation of Commitment", and "Complexity". Eight items loaded on the thirteenth factor, labelled "Teamwork", and four items loaded on the fourteenth factor labelled "Resource Management". Table 1 displays the mapping of the initial 19 CF factors to the validated 14 factors.

The second order EFA was conducted on the 14 factors derived from the first order EFA output. The PAF extraction method and the Oblimin with Kaiser Normalization (oblique) rotation method were used to calculate the scores. The second order EFA generated the KMO measure of sampling adequacy test result of 0.779 and a Bartlett’s test for Sphericity significance level of 0.00 which made it viable to conduct an EFA. The Eigenvalues generated from the PAF extraction method resulted in 4 constructs, with the Eigenvalues greater than or near 1 and the Scree plot identifying the valid constructs. The cumulative percentage explained by the four constructs is 67.8%.

In summary the second order EFA was applied to the 14 factors calculated in the first order EFA. The PAF method was used to extract the factors, followed by the Oblimin with Kaiser Normalization (oblique) rotation method. Of the 14 input factors, only four factors were retained for rotation, because of their Eigenvalue being

greater than or near one. The first four factors as a collective accounted for 67.8% of the cumulative variance. These four factors are consequently referred to as the four constructs of the SACCF.

*Table 1: Mapping of the initial 19 factors to the validated 14 factors.*

<b>Fourteen Factors Loaded from Questionnaire Items</b>	<b>Nineteen Factors based on Literature Review</b>
Organizational Behavior	<ul style="list-style-type: none"> <li>➤ Organizational Structure</li> <li>➤ Management Support</li> <li>➤ Organizational Culture</li> </ul>
Sprint Management	<ul style="list-style-type: none"> <li>➤ Sprint Management</li> <li>➤ Change Resistance</li> </ul>
Relative Advantage	<ul style="list-style-type: none"> <li>➤ Relative Advantage</li> </ul>
Experience	<ul style="list-style-type: none"> <li>➤ Experience</li> </ul>
Training	<ul style="list-style-type: none"> <li>➤ Training</li> </ul>
Specialization	<ul style="list-style-type: none"> <li>➤ Specialization</li> </ul>
Recognition	<ul style="list-style-type: none"> <li>➤ Recognition</li> </ul>
Customer Collaboration	<ul style="list-style-type: none"> <li>➤ Collaboration</li> <li>➤ Quality</li> </ul>
Compatibility	<ul style="list-style-type: none"> <li>➤ Compatibility</li> </ul>
Over-Engineering	<ul style="list-style-type: none"> <li>➤ Over-Engineering</li> </ul>
Escalation of Commitment	<ul style="list-style-type: none"> <li>➤ Escalation of Commitment</li> </ul>
Complexity	<ul style="list-style-type: none"> <li>➤ Complexity</li> </ul>
Teamwork	<ul style="list-style-type: none"> <li>➤ Teamwork</li> <li>➤ Communication</li> </ul>
Resource Management	<ul style="list-style-type: none"> <li>➤ Resources</li> </ul>

#### IV. RESULTS

The previous section described the methodology used to derive to the validated factors and constructs of the Conceptual Framework (CF). A statistical analysis of the results derived with this methodology, is presented in this section.

##### **a) Testing the Fourteen First Order Factor Relationship Strength**

A correlation matrix was used to test for the relationship strength among the different factors. A Spearman correlation analysis was conducted on all the factors as opposed to a Pearson correlation analysis, due to the skewness of the data discovered during the normality tests. The Spearman correlation analysis

revealed statistically significant correlations for the relationships between Scrum Adoption and all the factors at the 0.01 level, except for Teamwork which was significant at the 0.05 level ( $p=0.018$ ), and Over-Engineering with no significance ( $p=0.514$ ), see Table 2.

##### **b) Testing the Four Second Order Factor Relationship Strength**

A correlation matrix was used to test the relationship strength among the four constructs, as well as between the four constructs and the dependent variable. A Spearman correlation analysis was conducted as opposed to a Pearson correlation analysis, due to the skewness of the data discovered during the normality tests. Spearman correlation analysis revealed statistically significant correlations for the relationships between Scrum Adoption and the four constructs at the 0.01 level, see Table 3.

##### **c) Testing the Statistical Significance of the Factor Relationship**

All the normality assumptions were met when a regression analysis was conducted on the 14 factors. Tolerance values were above .01, and all the VIF values were below 10, and the assumption of no multicollinearity was met. The Durbin-Watson statistic fell within an expected range, which suggests that the assumption of no autocorrelation of residuals was met. The assumptions of linearity and homoscedasticity were also met, since the Scatterplot of standardized residual and standardized predicted value did not curve or funnel out. The normal probability plot of the residuals was approximately linear, which suggests that the assumption of normality of residuals was also met.

For the 14 factors, Multiple Linear Regression (MLR) was conducted to examine whether Over-Engineering, Relative Advantage, Recognition, Experience, Teamwork, Specialization, Escalation of Commitment, Compatibility, Resource Management, Customer Collaboration, Complexity, Training, Sprint Management, and Organizational Behavior impact on Scrum Adoption. The overall model (predictors: Over-Engineering, Relative Advantage, Recognition, Experience, Teamwork, Specialization, Escalation of Commitment, Compatibility, Resource Management, Customer Collaboration, Complexity, Training, Sprint Management, Organizational Behavior) explained 52.9% of the variance of Scrum Adoption, which was revealed to be statistically significant ( $F(14,206)=15.40$ ,  $p<0.0001$ ).

Table 2: Correlations among all the Factors used in the study.

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15
F1	1.00	.30**	.28**	.30**	.66**	.22**	.23**	.20**	.34**	.50**	.22**	.34**	.16*	.20**	.05
F2	.30**	1.00	.14*	.32**	.29**	.26**	.25**	.19**	.20**	.23**	.27**	.19**	.21**	.06	.09
F3	.28**	.14*	1.00	.25**	.29**	.58**	.24**	.66**	.72**	.27**	.30**	.36**	.16*	.64**	-.18*
F4	.30**	.32**	.25**	1.00	.10	.25**	.01	.09	.26**	.09	.08	.10	.71**	.16*	.26**
F5	.66**	.29**	.29**	.10	1.00	.29**	.27**	.24**	.35**	.64**	.28**	.51**	.01	.24**	-.02
F6	.22**	.26**	.58**	.25**	.29**	1.00	.28**	.65**	.51**	.23**	.21**	.26**	.10	.39**	-.01
F7	.23**	.25**	.24**	-.01	.27**	.28**	1.00	.24**	.31**	.32**	.34**	.31**	-.07	.24**	-.23**
F8	.20**	.19**	.66**	.09	.24**	.65**	.24**	1.00	.55**	.24**	.16*	.34**	.07	.48**	-.09
F9	.34**	.20**	.72**	.26**	.35**	.51**	.31**	.55**	1.00	.29**	.29**	.39**	.11	.57**	-.12
F10	.50**	.23**	.27**	.09	.64**	.23**	.32**	.24**	.29**	1.00	.22**	.58**	.01	.25**	-.04
F11	.22**	.27**	.30**	.08	.28**	.21**	.34**	.16*	.29**	.22**	1.00	.27**	-.02	.30**	-.33**
F12	.34**	.19**	.36**	.10	.51**	.26**	.31**	.34**	.39**	.58**	.27**	1.00	.01	.42**	-.14*
F13	.16*	.21**	.16*	.71**	.01	.10	-.07	.07	.11	.01	-.02	.01	1.00	.13	.28**
F14	.20**	.06	.64**	.16*	.24**	.39**	.24**	.48**	.57**	.25**	.30**	.42**	.13	1.00	-.24**
F15	.05	.09	-.18*	.26**	-.02	-.01	-.23**	-.09	-.12	-.04	-.33**	-.14*	.28**	-.24**	1.00

F1=Scrum Adoption, F2=Experience, F3=Organizational Behavior, F4=Sprint Management, F5=Relative Advantage, F6=Training, F7=Specialization, F8=Recognition, F9=Customer Collaboration, F10=Compatibility, F11=Escalation of Commitment, F12=Complexity, F13=Teamwork, F14=Resource Management, F15=Over-Engineering.

N Missing 0

\*\* . Correlation is significant at the 0.01 level (2-tailed).

\* . Correlation is significant at the 0.05 level (2-tailed).

Table 3: Correlations between the Four Constructs and Scrum Adoption.

	Scrum Adoption	Individual	Organization	Team	Technology
Scrum Adoption	1.00	.29**	.30**	.20**	.53**
Individual <sup>1</sup>	.29**	1.00	.39**	.16*	.38**
Organization	.30**	.39**	1.00	.25**	.42**
Team <sup>1</sup>	.20**	.16*	.25**	1.00	.07
Technology	.53**	.38**	.42**	.07	1.00

N Missing 0

\*\* Correlation is significant at the 0.01 level (2-tailed).

\* Correlation is significant at the 0.05 level (2-tailed).

<sup>1</sup>=factor's negatively phrased questions were recoded.

An inspection of the individual predictors of the overall model revealed that Relative Advantage (Beta=0.688,  $p < 0.0001$ ), Sprint Management (Beta=0.109,  $p < 0.05$ ), and Complexity (Beta=0.041,  $p < 0.05$ ) are significant predictors of Scrum Adoption (Table 4). Higher levels of Relative Advantage are associated with higher levels of Scrum Adoption, higher levels of Sprint Management are associated with higher levels of Scrum Adoption, and higher levels of Complexity are associated with lower levels of Scrum Adoption.

For the four constructs, MLR was conducted to examine whether Individual Factors, Technology Factors, Team Factors, and Organization Factors impact on Scrum Adoption. The overall model explained 33.40% of the variance in Scrum Adoption, which was revealed to be statistically significant ( $F(4,206)=25.34$ ,  $p < 0.0001$ ). An inspection of the individual predictors revealed that Technology Factors (Beta=0.580,  $p < 0.0001$ ) and Team Factors (Beta=0.126,  $p < 0.05$ ) are significant predictors of Scrum Adoption (see Table 5). Higher levels of Technology Factors are associated with higher levels of Scrum Adoption, and higher levels of Team Factors are associated with higher levels of Scrum Adoption.

## V. DISCUSSION OF FINDINGS

It is important to note that initially, the Scrum Adoption Challenges Conceptual Framework (SACCF) had 19 factors (independent variables). However, during the validation of the scale, the Exploratory Factor Analysis (EFA) applied to the questionnaire items extracted 14 factors. The loading of the questionnaire items to new factors meant that the initial predicted model had to be evaluated. The questionnaire items with its commonalities and corresponding factor loadings were studied and it was found that the initial 19 independent variables loaded correctly into the 14 factors. The new factor loadings, therefore, made logical sense. In Table 1, as discussed in

Section III, the 19 hypothesized factors are mapped to the newly validated 14 factors.

While most of the mappings in Table 1 is self-explanatory, it is necessary to give an explanation of the four factors that have more than one variable. These four factors are:

- Organizational Behavior
- Sprint Management
- Customer Collaboration
- Teamwork

The term Organization Behavior (OB) is defined as the actions and attitudes of individuals that work within an organization. OB is, therefore, the study of human behavior within the organizational environment, how human behavior interacts with the organization, and the organization itself [5]. George et al. [5], also states that the manner in which managers manage others is significantly affected by OB. Given this perspective of OB, it is reasonable to load Organizational Structure, Management Support, and Organizational Culture as a single factor under the heading OB.

The loading of Sprint Management and Change Resistance into a single factor is also logically sensible since firstly, Sprint Management is a time-boxed activity. Scrum practitioners would be performing their tasks within a Scrum sprint under most circumstances although it is recognized that this may not be the case for every task performed. Consequently, if a team is resisting change, it would manifest when the change is requested or performed during the Scrum sprint. To reiterate the fourth value of Agile development, which is "responding to change over following a plan", it is therefore fitting that Sprint Management and Change Resistance loaded as the Sprint Management factor, since Change Resistance by default, is part of the Sprint Management cycle [6].

Table 4: Regression Coefficients of the 14 Factors.

Coefficients <sup>a</sup>						
Model		Unstandardized Coefficients		Standardized Coefficients Beta	t	Sig.
		B	Std. Error			
1	(Constant)	.506	.454		1.114	.267
	Experience	-.021	.051	-.026	-.419	.676
	Organizational Behavior	.000	.062	.000	.003	.998
	Sprint Management <sup>1</sup>	.109	.049	.178	2.239	.026
	Relative Advantage	.688	.068	.702	10.168	.000
	Training	-.031	.052	-.045	-.604	.547
	Specialization	.004	.042	.006	.103	.918
	Recognition	-.019	.047	-.032	-.410	.682
	Customer Collaboration	.118	.062	.151	1.900	.059
	Compatibility	.085	.058	.099	1.477	.141
	Escalation of Commitment	.011	.041	.018	.280	.780
	Complexity	-.116	.056	-.146	-2.061	.041
	Teamwork <sup>1</sup>	-.013	.047	-.021	-.279	.781
	Resource Management	-.042	.051	-.059	-.830	.407
Over-Engineering <sup>1</sup>	.004	.039	.005	.092	.927	

<sup>a</sup>. Dependent Variable: Scrum Adoption

<sup>1</sup>=factor's negatively phrased questions were recoded.

Table 5: Regression Coefficients of the 4 Constructs.

Coefficients <sup>a</sup>						
Model		Unstandardized Coefficients		Standardized Coefficients Beta	t	Sig.
		B	Std. Error			
1	(Constant)	1.197	.445		2.692	.008
	Team <sup>1</sup>	.126	.062	.123	2.040	.043
	Technology	.580	.064	.566	9.009	.000
	Individual <sup>1</sup>	.016	.053	.019	.303	.763
	Organization	-.033	.054	-.039	-.616	.539

<sup>a</sup>. Dependent Variable: Scrum Adoption

<sup>1</sup>=factor's negatively phrased questions were recoded.

The loading of Collaboration and Quality into the Customer Collaboration factor was easy to accept since Customer Collaboration entails working closely with the client in order to deliver what was requested at the expected quality. The last merged factor loading was Teamwork which consists of Teamwork and Communication. This factor loading was also a simple decision and with hindsight, these two factors had to be grouped together from the outset. The reason for this is because Teamwork requires individuals to work together to complete tasks, and communication is a critical component to complete sprint tasks within the team. It is important to note that the Resources factor has been renamed to Resource Management because resource shortage or surplus is a management related concern.

Figure 2 displays the third and final iteration of the CF. The hypothesized relationships between the independent variables and the dependent variable are shown in the parenthesis. As is evident from the diagram, the conceptual model is much more refined than the previous iterations. The Specialization factor which was previously under the team construct is now under the individual construct, and Over-Engineering which was an individual factor is now a team factor. The reason for these realignments is because Specialization or specialized skills can be narrowed down to the individual level. Over-Engineering, if encountered and allowed within a Scrum team environment, means that the team was not vigilant enough during their communication sessions to identify

when an individual was doing more than what was required.

Four of the initial 19 factors were revealed as having a significant linear relationship with Scrum adoption. The four factors are Relative Advantage, Complexity, Change Resistance, and Sprint Management. The factor that came close to having a significant relationship with Scrum adoption was Customer Collaboration with  $p=0.059$ . Because of the new factor loadings Sprint Management and Change Resistance loaded onto Sprint Management, as noted earlier.

## VI. CONCLUSION

Scrum and Agile software development, including Scrum adoption, is a growing phenomenon. The research presented in this paper contributes both towards the Agile body of knowledge and to Scrum adoption. A proposed consolidation of Scrum and Agile challenges, a Conceptual Framework (CF), and the evaluation of the CF using quantitative methods and techniques were explored in this paper. The primary objective of this paper was the investigation of factors that have a significant linear relationship with Scrum adoption as perceived by Scrum practitioners working within SA organizations. Three validated factors which have a significant linear relationship with Scrum adoption have been identified.

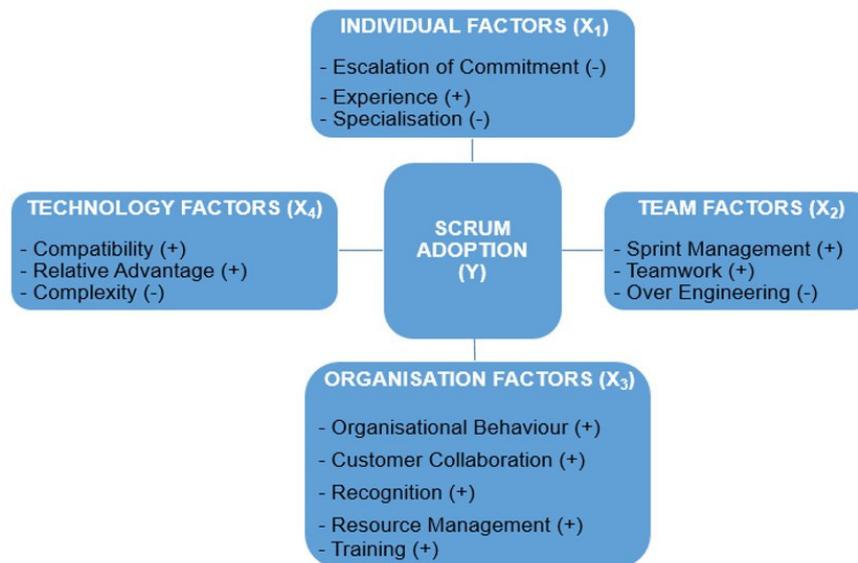


Figure 2: Final Iteration of the Conceptual Framework.

This research can be extended by a systematic review of existing Scrum and Agile adoption challenges, as well as a larger population sample for greater generalization of the findings. For future research it would be beneficial to develop a logistic regression model for predicting an organization's success rate at Scrum adoption based on the organization's current practices. The predictive analysis can be conducted by comparing the test data of the organization to the trained data model derived from the population sample.

#### REFERENCES

- [1] Overhage, S., Schlauderer, S., Birkmeier, D. & Miller, J. 2011. What Makes IT Personnel Adopt Scrum? A Framework of Drivers and Inhibitors to Developer Acceptance. In 2011 44th Hawaii International Conference on System Sciences. IEEE: 1–10. <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5718964>.
- [2] Chan, K.Y. & Thong, J.Y.L. 2007. An Integrated Framework of Individual Acceptance of Agile Methodologies. PACIS 2007 Proceedings: 154.
- [3] Hanslo, R. & Mnkandla, E. 2018. Scrum Adoption Challenges Detection Model: SACDM. In Federated Conference on Computer Science and Information Systems (FedCSIS). Poznan, Poland: IEEE: 949–957.
- [4] Gerber, H. & Hall, R. 2016. Statistical analysis reporting template for researchers.
- [5] George, J.M., Jones, G.R. & Sharbrough, W.C. 2005. Understanding and managing organizational behavior. Upper Saddle River, NJ: Pearson Prentice Hall.
- [6] Beck, K., Beedle, M., Van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., Grenning, J., Highsmith, J., Hunt, A., Jeffries, R., Kern, J., Marick, B., Martin, R.C., Mellor, S., Schwaber, K., Sutherland, J. & Thomas, D. 2001. Agile Manifesto. Software Development, 9. <http://agilemanifesto.org/>.
- [7] VersionOne. 2015. 9th Annual State of Agile Survey. <http://stateofagile.versionone.com/>.
- [8] Schwaber, K. & Sutherland, J. 2011. The scrum guide. Scrum.org, October, 2: 17. <https://www.scrum.org/index.php/resources/scrum-guide>.
- [9] Mohan, K. & Ahlemann, F. 2013. Understanding acceptance of information system development and management methodologies by actual users: A review and assessment of existing literature. International Journal of Information Management, 33(5): 831–839.
- [10] Noruwana, N. & Tanner, M. 2012. Understanding the structured processes followed by organisations prior to engaging in agile processes: A South African Perspective. SACJ, (48): 8.
- [11] Chan, F.K.Y. & Thong, J.Y.L. 2009. Acceptance of agile methodologies: A critical review and conceptual framework. Decision support systems, 46(4): 803–814.
- [12] Sultan, F. & Chan, L. 2000. The adoption of new technology: the case of object-oriented computing in software companies. IEEE transactions on Engineering Management, 47(1):106–126.
- [13] Rogers, E.M. 2003. Diffusion of Innovations, 5th Edition. Free Press. <https://books.google.co.za/books?id=9UIK5LjUOwEC>.
- [14] Kishore, R. & McLean, E.R. 2007. Reconceptualizing innovation compatibility as organizational alignment in secondary IT adoption contexts: an investigation of software reuse infusion. IEEE Transactions on Engineering Management, 54(4): 756–775.
- [15] Derish, P.A. and Annesley, T.M., 2011. How to write a rave review. Clinical Chemistry, 57(3), pp.388-391.
- [16] Welman, C., Kruger, F. & Mitchell, B. 2005. Research Methodology. 3rd ed. Cape Town, South Africa: Oxford University Press.
- [17] Dingsøyr, T., Hanssen, G.K., Dybå, T., Anker, G. & Nygaard, J.O. 2006. Developing Software with Scrum in a Small Cross-Organizational Project. In I. Richardson, P. Runeson, & R. Messnarz, eds. Software Process Improvement. Joensuu: Springer: 5–15. [http://link.springer.com/10.1007/11908562\\_2](http://link.springer.com/10.1007/11908562_2).



# Create your own agile methodology for your research and development team

Enikő Ilyés

Eötvös Loránd University  
in Budapest

sny. Pázmány Péter 1/C, 1117 Budapest, Hungary  
Email: ilyese@inf.elte.hu

**Abstract**—Agile methodologies conquer space beyond their industrial use. For applying them in a situation other than classic software development, one should first assess the features of this specific environment. As a next step, elements from various well-known agile methodologies (roles, events, products) can be considered as building blocks. These elements combined in a new way, adopted appropriately result in a specific, own agile methodology. In this study, we present a list of aspects that one should consider when creating a specific agile methodology for a R&D team. Our own agile methodology created for the txtUML R&D team from the Faculty of Informatics of Eötvös Loránd University is built along this list of aspects. Known and new agile elements were included in this specific methodology and are explained in this article in detail. The txtUML R&D methodology has been used with satisfaction since 2018, as evidenced by backward surveys.

## I. INTRODUCTION

THE introduction of agile methodologies has been of great interest in recent years in Central Europe [9], [10], [11]. They are used not only in software engineering, but also in other fields - such as economics-, and not only within the industry, but in research and development as well [1], [2], [3], [4], [6]. Naturally, every time agile methodologies are applied in a new area, the original agile framework needs to be adapted to the specifics of the new field.

In 2018 we decided to develop a specific agile methodology to manage the work of a research and development team at the Eötvös Loránd University. In the research and development teams of our university teachers, research staff and students work together. Students can be PhD students, master's students or undergraduates. The diversity of the team members influences the frequency of meetings. The fluctuation of team members results in different levels of knowledge, differences in the ability to work independently, as well as varied levels of motivation. Therefore, many aspects of collaboration need to be addressed in order to ensure the efficient operation of these teams.

In the second chapter of this paper we summarize some examples of agile methodology used within the research and development teams of other universities. In the third chapter, we present a list of aspects we used and recommend to be taken into consideration before creating a new methodology

The research project was supported by the European Union and co-financed by the European Social Fund (EFOP-3.6.3-VEKOP-16-2017-00002).

for a research and development team. In the fourth chapter we present our own methodology, along with the related feedback. Finally, the fifth chapter summarizes the results.

## II. AGILE METHODS IN R&D TEAMS

If we look at the characteristics of agile methodologies, we can see that they fit properly with the nature of research and development teams. Maik Sayfert, who has more than 10 years real project experience with agile software development, declares: “When considering research as an area of high uncertainty and open results, it smells like agile methods are perfectly suited. From all methods I personally can imagine Scrum with short/mid sized iterations it the best match, as Kanban (flow based) is more suited to connect several disciplines/departments and XP is more technically driven.” [15]

Jeff Sutherland, the inventor and Co-Creator of Scrum reports in 2016: “Many of the leading research labs in the U.S. use Scrum. The one I have worked with most often is the John’s Hopkins Applied Physics Lab, the leading Naval research lab. Their research plan is their backlog. They map it out like an AI tree. Time boxing the research stories gets them done twice as fast. And the quality of the research is much higher with daily meetings.” [16]

Some articles present concrete cases using agile methodology to manage a research and development team.

There are several cases where the introduction of agile methodology is related to the will of an industry partner, who uses agile methodologies from earlier. For the agile coordination of the research team and the industrial partner, traditional roles may be modified or extended. For example H. Sharp and co-authors in [1] reports such a case: “an important adaptation was the inversion of the semantic of the product owner, for in our context he is a member of the lab allocated at the client. This change was made because of the difficulties associated with having a client in the lab.” Some research teams have developed new agile methodologies for improving collaboration with industrial partner, such as the the Agile Research Network collaboration model. The creators of this model found the following key challenges in this kind of collaboration: timeliness, relevance, rigour, access. Based on this recognition they build the ARN model, which consists of the following sections: Collaboration Kick-off, Investigation of the focus area, Implementation, Evaluation. [2]

The introduction of agile methodologies in research and development teams are not necessarily motivated by cooperation with industry but rather by coping with other challenges, such as distributed teams. In such cases, new roles other than the original Scrum roles are introduced (eg. Unit Coordinator, Research unit, etc.). Events are mostly online, such as: Daily Scrum is done via instant messaging. [3]

The inventors of the SCORE agile method for research and development teams addressed another challenge: “Working with and mentoring Ph.D. students is the central activity in running an academic research group, with two broad goals: (1) to collaboratively produce high-quality research results, and (2) to help students to become independent researchers capable of working at research labs or academic institutions. At first, we followed a simple, fairly typical approach to mentoring: we met once or twice per week with each student in roughly half-hour or hour-long slots. Unfortunately, as the number students grew from two or three each to six or seven each, and as our outside commitments steadily increased, our simple approach reached its limits.” [4]

SCORE methodology uses elements such as status meetings (daily Scrum), on demand technical meetings, weekly reading group, weekly lunch; The core idea is to keep status separate from research. “The Scrum meeting is for status, and the on-demand meetings are for solving unforeseen problems. Keeping the two activities separate allows them to be undertaken more efficiently.” [4]

Based on the idea of “the last step of having a successful team is to build a supportive environment.” the authors of [5] proposed an approach for planning research projects, considering the use of 2-type fuzzy numbers for research project planning based on Scrum.

Many positive experiences were drafted regarding using agile methods in research and development teams. Some examples: “using Scrum contributed to the education of team members that were interns (students) and/or autonomous professionals (freelancers)” [6] “The involvement and commitment of members of the team with the results increased (...). We also realized that team members were motivated and open to changes in work.” [1] “Everything is visible to everyone, team communication improves, a culture is created where everyone expects the project to succeed.” [3] “Students say they are more productive, more enthusiastic about research, and have better interactions with other students and with their adviser. Students reported that there is now a real sense of community in the group that was never there before.” [4] “Though SCORE is conceptually simple, its benefits to us have been significant. (...) when students are struggling, it often only takes a day or two to realize something is not right, and to begin to address it. Our time is spent far more effectively.” [4]

The text-based feedbacks sound good, but unfortunately it is still a challenge to find a metric to evaluate formally the gain achieved by using agile methodologies in R&D teams [1]. Another claim is that university students do generally not have the maturity to understand the agile practices and their consequences. “The suitability of agile methods in education

is thus an ongoing debate.” [7]

### III. ASPECTS TO BE ANALYZED WHILE CREATING A SPECIFIC AGILE METHOD FOR A R&D TEAM

Based on literature and our experiences, we have gathered a list of aspects to take into consideration if someone wants to create an appropriate agile methodology for their research and development team.

- 1) Objectives of the research and development team: As we have mentioned, a research and development team may have several goals with varying degrees of emphasis. The team’s primary goal may be to develop a new product, but another goal may be to involve students more effectively in the research, or even to recruit researchers and teachers. These goals need to be clarified, prioritized and then tailored to the research team’s workstyle.
- 2) Proportion of research and development tasks: The new methodology should sufficiently support team members in both research and development tasks.
- 3) Type of members: The group may include undergraduate, master’s and PhD students, as well as research staff and teachers. These individuals have varying levels of prior knowledge, may differ in terms of their ability to work independently, and may be able to devote different amounts of time to the R&D team’s work.
- 4) Motivation of team members: Team members participate in the R&D work with different motivations. Some students volunteer on the team to get to know the world of research, while others have more specific goals, such as writing their thesis. There are also students who work for university credits and others to receive a scholarship. Research staff is mostly motivated by their interest in the topic of research and possible results. Teachers are primarily motivated by recruiting teachers and researchers. When developing a unique methodology, all these motivations need to be taken into account to create an engaged team.
- 5) Fluctuation of team members: It is important to consider turnover within a team. For example, while an undergraduate student stays on the team for an average of half a year, a PhD student is likely to remain a team member for several years. New team members have to be integrated into the team as early as possible, and keeping existing knowledge within the team when a significant member leaves is also very important.
- 6) Number of team members: The number of people who have to work together has always been a very important aspect of managing a team. Large groups can benefit from a greater knowledge base, but communication can become more difficult. The efficiency of events and different roles assigned is highly influenced by the number of people on the team.
- 7) Frequency of meetings: As students, teachers and researchers perform many different tasks at the university during working hours, it is not easy to organize team meetings. The team’s new working methodology will be

strongly influenced by the possible number and timing of meetings.

- 8) Relation between tasks: It is recommended to take into consideration how the potential research and development tasks are built on each other and how connected they are to each other. For example, if some tasks have the same topic, it is recommended that team members taking on these responsibilities work together and share their experience. Perhaps the creation of subgroups along task topics would be effective, including the appointment of a subgroup leader.
- 9) Documentation obligations: The level of detail needed in documenting the group's activities is an important factor. For example, if a company financially supports the research and development team, they may have to meet specific documentation requirements. Another purpose of detailed documentation could be to attenuate the effects of high turnover among team members; the group needs to ensure that the knowledge of its members remains in the team even after they leave. On the other hand, if the team does not need documentation for any purpose, creating a detailed record may be a waste of the members's valuable time.
- 10) External partners: Beyond the members of the research team, there may be a sponsor, a customer or a partner whose expectations they need to meet, such as performing a certain task by a certain deadline. Working with an external partner requires special attention: Who is the partner? What deadlines have they set for the team? How can the team achieve the fulfillment of the partner's requirements without sacrificing its own goals?

Please note, that the list is not exclusive. For example, there is a growing phenomenon that the members of the research team are not co-located, which poses additional challenges while creating a new methodology

#### IV. OUR SPECIFIC AGILE METHOD

##### A. The *txtUML* R&D team

The Faculty of Informatics at the Eötvös Loránd University currently has several functioning research and development labs. The Model Driven Development Research Lab has been operating since 2014, with a project called *txtUML* as one of its primary undertakings. A brief description of the software is: "The name *txtUML* stands for textual, executable and translatable UML. It is an open source project with the goal to make model driven development easier." (For a more detailed description see the website of *txtUML* [12]. Before we have created the agile methodology for the *txtUML* research team, I attended their meetings for half a year. I considered that the work of the team was not efficiently organized. The meetings consisted of telling everyone what they were doing last week and what problems they encountered. When someone came in, he/she talked about his/hers problems for a long time. One or two people who understood the topic discussed technical details, while others waited quietly. This took 2 hours, so the

meetings were long and demotivating. Students who joined the team from the beginning of the semester could not involve in the discussions till the middle of the semester. There was no time allocated for them to learn from the experienced students during the regular meetings. This would have been a great necessity and possibility, since the students in the team had different level of understanding of the subject, and there were also undergraduates, masters and doctoral students in the team. Many types of motivation were present in the team thanks to the many types of students - some of them joined the research group as a course named "Software Technology Lab.", others wrote their theses, etc. I saw that the methodology does not take into account the presence of different motivations and, as a result, there is not enough commitment to work. The research team was supported by a tender that required weekly logging and fulfillment of predefined goals. I considered that this methodology did not help students enough to fulfill this expectation. As a result of my observations I collected all the important factors which made the actual method work ineffectively. This list was the basis of the list I presented in the previous chapter.

In 2018 we decided to introduce an agile methodology to manage the *txtUML* team's work more efficiently. Taking into consideration the particularities of our team, designing our own agile methodology using elements from other agile methods as building blocks and inspiration seemed to be the best solution.

The particularities of our team were:

- 1) Objectives of the research and development team: The development of *txtUML* software and the involvement of students in the world of research are equally important objectives for us.
- 2) Proportion of research and development tasks: We predominantly perform development tasks (80%), with research tasks being less frequent (20%).
- 3) Type of members: Our team is currently comprised of 9 undergraduate students, 1 PhD student and 1 teacher. The individuals have different levels of prior knowledge regarding *txtUML*: three senior students who have been working on our team for more than a year; three students who joined the team six months ago and three completely new students.
- 4) Motivation of team members: The 9 undergraduate students gain 4 credits/semester for their work on the team, while 4 of the students also receive a scholarship as a result of participating in this research. The PhD student improves her PhD work with the help of this team, and the one teacher's primary motivation is the recruitment of new teachers and researchers.
- 5) Fluctuation of team members: 3 students are expected to leave the team at the end of the semester, while others may stay for 2 more semesters.
- 6) Number of team members: 10
- 7) Frequency of meetings: A 2-hour meeting/week is attainable.
- 8) Relation between tasks: The research and development

tasks of our team can be categorized in four groups. The following terms are used to refer to their topic: language front-end, visualization, C ++ export and model testing.

- 9) Documentation obligations: Due to the scholarship some student receive in relation to their work with txtUML, we must log the progress of the team members on a weekly basis. As three senior students leave the team at the end of the semester, we need to ensure that their knowledge is “saved” (documented) before they leave the group.
- 10) External partners: Currently, our only external partner is the scholarship program. At the beginning of each semester, we have to make commitments and present out progress at the end of the semester.

As a member of the txtUML team with a “team coach” role, I design a new agile method for the txtUML R&D team based on the team's characteristics. The methodology is presented below.

### *B. txtUML's agile methodology*

In the following, I will describe the agile methodology designed for the txtUML team and used by it since February 2018. I will group the elements of the methodology along roles, events and artifacts. Regarding each element, I will clarify the purpose for which it was introduced and between {} signs I will specify which aspects from the list “Aspects to be analyzed while creating a specific agile method for a R&D team” (Chapter 4.) are mostly related to the mission of that specific element.

1) *Roles*: There are five different roles in the txtUML team: project leader, Scrum master, technical leader, subgroup leader, developer.

**Project leader**: Represents the goals of the research team towards the university leadership and applications. He is the main contact person of the team, holds up the results of the team (for example, in the case of “University Open Day”, “Researchers’ Night” events). He has a word in any decision-making. His role provides a solid framework and direction for the research team. He participates on group meetings and is also easily accessible between them. { 1, 4, 9, 10 }

**Scrum master**: The name of the role was inspired by the Scrum methodology. Similarly, the responsibility of this person is to protect the team’s operational values. This can be achieved by initiating and coordinating of various events (for example, initiating and coordinating weekly Scrum for effective discussions; training the members to effectively report their blockers, etc.) and leading by example. The Scrum master participates on group meetings. She monitors the situation of each team member and the dynamics of the whole group. She does not have software development tasks. { 3, 4, 5, 6, 9 }

**Technical leader**: Has a comprehensive view of the software and is also familiar with many details. Coordinates team members’ work from a technical point of view. All important technical decisions must be agreed with him. It also has a mentoring role: to pass on the knowledge and experience of the previous developers. He is primarily available at group

meetings, but often responds to questions through online communication as well. { 1, 2, 3, 4, 5, 6, 7 }

**Subgroup leader**: The research and development tasks of our team can be distinguished in four groups. The following terms are used to refer to their topic: language front-end (3 members); visualization (2 members); C ++ export (2 members); model testing (2 members). Students working on the same topics form a subgroup led by a senior student who has the most knowledge and experience in that specific field. This student designates smaller research and development issues to subgroup team members and gives developers mentoring during meetings and through online communication. Usually he/she performs development tasks as well. During meetings he/she organizes discussions on its own topic to transfer its knowledge and presents the results of its subgroup to the entire research group. { 1, 3, 4, 5, 6, 7, 8 }

**Developer**: As a first step in joining the team, every new member gets acquainted with txtUML from the “txtUML user” perspective. Than he/she selects the area in which he wishes to contribute to the research and development work and thus becomes a member of a subgroup. With the help of the subgroup leader he/she chooses an appropriate task and contributes to the expansion and refinement of the txtUML software. Developers also participate on group meetings where they learn from others and pass on their experience as well. { 1, 3, 4, 5 }

Remark: It may be strange that the role of the Product Owner is not strongly emphasized, though it is a key role in the Scrum methodology for example. According to my observation, in the university environment, the role of Product Owner is often less powerful, perhaps because the role of the customer is not as definable as in the industrial environment. Most often, the teacher fills the role of the customer. For him/her, however, the functionality of the product is not the only focus, but there are also pedagogical goals. In our case, the role of the Project leader was as close as possible to the Product Owner: he pointed out the development directions; the developers, subgroup leaders talked about the functions to be realized and their acceptance criteria with him.

2) *Events*: The members of the txtUML team meet weekly in a two-hour session. Events related to the agile methodology are integrated into this meeting. In the intervening period everyone works individually on their own task and uses the slack ( [13]) tool for online consultation, if needed.

Our four major agile events are: preparation, weekly routine, retrospective, demonstration.

**Preparation**: The preparation phase takes the first 2-3 meetings of the semester. The project leader, technical leader and subgroup leaders discuss the main research and development directions of the semester. The new team members get to know the system, the other developers indicate what specific tasks they would like to take. The Scrum master presents the values of the group's operation and the methods that the group will use during the semester to support putting these values into reality (Example: Values: effective team discussions, sustaining motivation, transfer of knowledge; One

method used to support these values: weekly Scrum) {1, 2, 3, 4, 5, 6, 8}

**Weekly routine:** After the preparation phase come the weekly meetings with the following structure: short news, weekly Scrum, topic of the week, discussion with the whole team, discussion in small groups.

In the *news section*, after the greeting of team members, the Scrum master announces to the community if someone is going to miss the meeting or is going to be late; gives a brief reflection on the weekly reports submitted by the team members; notifies the group about the short news related to the R&D lab (For example: “We were invited to attend the University Open Day”). The *news section* grabs the focus of the team members and is an effective way to spread out the most important information regarding the team. {3, 6, 7, 10}

The *weekly Scrum* is a ritual with a timebox of 15 minutes. Technical leader, subgroup leaders and developers are required to participate. Each of them answers these three questions briefly: “What have I done since the last meeting?”; “What do I plan to do until the next meeting?”; “Is there anything that blocks me?”. There is a temptation to go into technical details that are important to one or two team members. It is important to avoid this, because if the other members aren't involved in that subject, they will become very bored and demotivated. If an important topic arises during the weekly Scrum the key words of that topic are recorded by the Scrum master and put on the table after the *weekly Scrum*, in the second half of the *weekly routine (discussion with the whole team, discussion in small groups)*. However, the weekly Scrum can include sharing of short tips with each other, as everyone is attentive to it. A positive side effect of the *weekly Scrum* is that as the team members speak out loudly about their rhythm of progress, they motivate themselves and each other as well. {1, 4, 6, 7}

The *topic of the week* is a 10-15 minutes period during which the Scrum master presents and trains a value of the team's operation. During the first half of 2018, there were topics such as: identifying individual motivation; identifying a common goal; understanding shared responsibility; formulating weekly reports in an effective way; estimating work left; etc. Discussion of such topics help new and old members to learn the values and methods that enable them to realize a true teamwork rather than working side by side. {1, 3, 4, 5, 6, 7, 10}

The *discussion with the whole team* and *discussion in small groups* phase provides an opportunity to discuss issues that have arisen during the weekly Scrum, but their extraction has been postponed to effectively assess the overall group situation anteriorly. For example, if a more serious technical decision must be made (that affects the overall structure of the product), the most involved members will present the question to be discussed and then everyone will argue about it. More of this kind of questions can come up at a weekly meeting, or not even one. If only a few people, or perhaps only one developer-and-subgroup leader-pair are involved in a question, then they separate and discuss it between themselves. An example of this kind of “pair cases” is when a developer gets stuck in his

own job and asks the subgroup leader for help; or is about to finish his/her task and asks for a new one. The *discussion section* provides an excellent opportunity for team members to share their experiences and develop their knowledge of a topic. {1, 2, 3, 4, 5, 6, 7, 8}

**Retrospective:** During the *retrospective* the team members can reflect on the quality of the team-work and formulate directions and methods that are likely to have a positive impact on it in the next iterations. In the first half of 2018, we held a retrospective during the seventh meeting of the semester. The retrospective was coordinated by the Scrum master, who handed out different-colored, small-sized paper sheets to each team member. One color stand for positive, the other color for negative feedback. On each sheet of paper, a key word or thought could be written by the team members in an anonymous way. The recollected sheets of paper were organized along the “negative / positive” and “technique / methodology / team” axes by the Scrum master on the board. She placed on top of each other the sheets of paper representing the same idea. The heaps appearing in this way warned: it would be worthwhile to deal with those feedbacks urgently, since they affect many team members. After reviewing all the feedbacks, a discussion took place during which the team formulated ideas for improving their team-work. A concrete example: more people have indicated that they perceive some administrative tasks as redundant - writing weekly reports, using GitHub [14], and participating on the weekly Scrum. During the meeting, we managed to clarify that each of them has separate goals. {1, 3, 4, 5, 6, 8}

**Demonstration:** The *demonstration* is the closing event of the semester, during which all members of the research and development lab summarize their work realized during the semester and present their results. Guests may also be invited to attend the event. The structure of the meeting is: the project leader presents the main research and development directions emphasized during the semester, the Scrum master gives information about the methodology used to organize the team work, and the developers, together with the subgroup leaders, demonstrate the implemented new functions. On July 2, 2018, this year's first demonstration took place, which was also a demonstration of the txtUML 0.7.0 release. On this occasion, old team members also joined us. Cake and champagne enhanced to the festive atmosphere, which aimed to emphasize the common success as motivation for upcoming team work. {1, 2, 3, 4, 5, 6, 10}

3) *Artifacts:* The primary product of the group's work is the developed code base, which is available on gitHub [14].

On the same website, you can find the tasks related to the software product, which can be considered as a *product backlog (under the issues tab)*. Developers, subgroup leaders, and technical leaders can add or select from these - of course, according to the priorities fixed by the Project leader. {1, 2, 6, 8, 9, 10}

Each of the subgroups has an *issue board* (also on the above-mentioned website, under the projects tab) that displays the current tasks of that subgroup. They include columns

like: 'to do'; 'in development'; 'under testing'; 'pull request'; 'done'. The board therefore provides a comprehensive picture of the current status of the subgroup's work. {1, 6, 7, 8, 9, 10}

In addition, there is an internal website, a form, used to collect information about the weekly progress of the members in a background table. It is called *weekly report*. The following data can be given on its surface regarding to a task: contributors name; type of the task (for example: research, development, mentoring, administration, etc.); a brief description of the task; a detailed description of the task (if the short description is not sufficient); associated GitHub task code (if any); lesson to be learned (if any). Developers must fill out the form every week before the meeting, for every task they worked on since the last meeting. In connection with the *weekly report*, the members of the team repeatedly expressed their disapproval. Nonetheless, we kept the weekly report because of its positive effect: since it precedes the *weekly Scrum*, it helps the group members to give a more focused report during the *weekly Scrum*; produces data series suitable for research and application accounting; can help in catching up for the team members that will join the team later on; collects data for self-reflection, logs personal performance; is a practice that develops the skills needed for group work (collaboration, organization, communication) and adapts to the administrative requirements of a general workplace. {1, 2, 5, 6, 7, 8, 9, 10}

4) *Bonus: Agile training*: In the second semester of 2018 I myself, as the agile coach of the team, organized an extra team event, called Agile training. The purpose of this was to consolidate agile values, strengthen team spirit, invent together new methods that improve team work.

The topics of the agile training were: getting know each other better - build trust, build relationships; clarifying the purpose and responsibilities of the different team roles; (re)defining the values of our team - taking inspiration from Scrum values and Google's research about effective team work; identifying the main motivation of each team member and the whole team (for that specific semester, regarding our research).

### C. Results and feedback

During the first semester of 2018, the methodology described above was applied for 18 weeks. Our observations were: weekly meetings were more efficient and dynamic than they were before; the number of communication interactions among team members has increased; the consciousness of group responsibility began to develop; regular administration of the tasks began to become accepted; individual and team goals related to the R&D lab have become clear.

At the review meeting (7th week), team members gave similar feedback: Teamwork was highlighted as a positive experience, even the collaboration between the subgroups. Shorter and useful meetings have been acknowledged. They were very positive about the senior students's competence, which, combined with the good meeting structure provided many learning opportunities for the new team members. The methodology used to provide the transfer of knowledge through group structure (hierarchical roles) and weekly routine

(mainly weekly Scrum and the discussion phase). Satisfaction with the effectiveness of the communication was also highlighted by more people. Three of the students reported that the collection and clear formulation of goals had a positive effect on them.

Negative feedback came primarily regarding the weekly report, which was indicated to be redundant. There was also a complaint regarding "over-formalizing the process". Our reaction was to give more time for these elements to prove their positive impact. If later reviews also highlight their demotivating effect, they can be omitted or modified following a common decision.

At week 17, team members filled out questionnaires regarding the whole semester. The questionnaires were filled by 7 people, i.e. 77% of the team. Some questions asked for a number as an answer, from a scale of 1 to 5, where "1" meant "the least", and "5" meant "the most". The question most relevant from the perspective of the methodology was: "To what extent do you think the value achieved during the semester is related to the project's management?" The average of the responses was 3.71. Another interesting question was "In what a measure did the values of the agile methodology prevail during the team work?" The average was a bit lower: 3.41. We don't consider these values very low, but we want to achieve a higher rating.

There were other aspects as well for that the rates received seemed to be considerably high. These could reflect the effect of the new agile methodology. These results are summarized in triplets below (question - average - possible explanation).

- To what extent was the task realized by you appropriate for you? - 4.28 - The *discussion in small groups* and the mentoring role of the subgroup-leaders helped team members to choose and succeed with a task appropriate for them.
- How satisfying was the number and quality of feedback you received regarding your work? - 4.57 - The *weekly Scrum*, the *discussion with the whole team* and *discussion in small groups*, the mentoring attitude of the subgroup leaders and the technical leader, provided space and possibility for giving individuals feedback.
- To what measure did the group work as a team? - 4.14 - The common reception of *news* at the beginning of the weekly meetings (for example, a joint congratulation to a team member success), the involvement of everyone in the *weekly Scrum*, the *topic of the week* helped to strengthen team spirit. The mentoring attitude of the technical leader and the subgroup leaders could also contribute to a sense of belonging to the team.
- How much did you enjoy being part of the team? - 4.71 - The *weekly routine* and the proper design of the roles helped to give every individual in the team enough attention and to work smoothly together.

According to some students, the best experiences were: "teamwork, development, encouraging each other"; "We have worked together on an interesting task." Concerning the "biggest challenges", we have noticed that team members who

had development and mentoring tasks as well struggled to allocate time for both of them. In the category of “What would you change?”, the fear of losing senior students (who complete their university studies) appeared, which was a warning to us that we should pay more attention to the transferring of the project knowledge. It appeared a need for deeper understanding of the methodology. (As a response to this need the above presented Agile training was introduced in the second semester of 2018, presented in chapter 5.)

For the results, please note that the research is still in an initial phase. The members of the group filled out the questionnaires, but since there are few of them, the feedback comes from a small number of people. Another lack of the research is, that team speed is not measured yet. Further iterations of the research are expected to improve regarding this aspects.

## V. SUMMARY

In order to manage the work of a software engineering research and development team operating at the university, we can get inspired by the agile methods used in the software development industry. There are some examples of how their use in a classical software development course has succeeded. However, if we want to use them in a research and development team, we have to keep in mind some characteristics like: the objectives of the research and development team, proportion of research and development tasks, type of members (student, teacher, researcher), documentation obligations, etc. A list of this aspects to be analyzed while creating a specific agile method for a R&D team is a result of this research.

Along these aspects, we developed an agile methodology for the txtUML research and development team. The *project leader's* guiding role assures well defined directions regarding the research, while the *Scrum master* assists for teamwork to run smoothly. *Technical leaders - subgroup leaders - developers* create a hierarchical chain targeting the efficient flow of the knowledge. Preparations, news section, topic of the week, and retrospective events are the key to raise awareness of team work values. *Weekly Scrum, product backlog, tables, and weekly reports* aim to enhance transparency and thus efficiency and motivation. The *discussion with the whole team* and *discussion in small groups* events assure knowledge transfer.

Regarding the usage of txtUML's agile methodology the feedback highlights: the communication between the team members has become more efficient; goals are clearer; the atmosphere is pleasant; teamwork is more effective; new members develop rapidly through the flow of the project knowledge. Feedback also highlighted further development opportunities, such as the introduction of on agile training for deepening agile values. All in all, we are satisfied with our methodology and we want to continue to “contribute to creating values for a group of people with passion and creativity, with the help of agile leadership” [8].

Mike Cohn, the famous Scrum trainer, the co-founder of the Scrum Alliance states “I hope we see an end to methodology

wars; Scrum vs. Kanban, SAFe vs. LeSS, Disciplined Agile, Enterprise Agile and every other scaling framework. Instead of arguing about methodologies, we need to focus more on agile as a large set of practices, some of which work well in combination.” [9]. We think that the approach presented in this article is a good example of how we can develop a suitable agile methodology for a team with specific characteristics by combining agile elements creatively, based on the analyses of some previously fixed aspects.

## ACKNOWLEDGMENT

I am grateful to all members of the txtUML R&D team who have welcomed all my ideas openly and helped to develop our own agile method.

The research project was supported by the European Union and co-financed by the European Social Fund (EFOP-3.6.3-VEKOP-16-2017-00002).

## REFERENCES

- [1] H. Sharp, L. Plonkas, K. Taylor and P. Gregory, “Overcoming challenges in collaboration between research and practice: the agile research network” *SER&IPs'14*, Hyderabad, India, 2014, pp. 10–13. doi:10.1145/2593850.2593859
- [2] L. Barroca, H. Sharp, D. Salah, K. Taylor and P. Gregory, “Bridging the gap between research and agile practice: an evolutionary model” *Springer*, 2015. doi: 10.1007/s13198-015-0355-5
- [3] M. Marchesi, K. Mannaro, S. Uras and M. Locci, “Distributed Scrum in Research Project Management” *Agile Processes in Software Engineering and Extreme Programming, 8th International Conference*, Como, Italy, 2007. doi: 10.1007/978-3-540-73101-6\_45
- [4] M. Hicks and J. S. Foster, “Adapting Scrum to Managing a Research Group” *Department of Computer Science Technical Report #CS-TR-4966*, 2010.
- [5] A. Klaus-Rosinska, J. Schneider and Vivian Bull, “Research Project Planning Based on SCRUM Framework and Type-2 Fuzzy Numbers” *ISAT 2018*, Nysa, Poland, 2018, pp. 381–391. doi: 10.1007/978-3-319-99993-7\_34
- [6] I. R. Lima, T. de C. Freire and H. A. X. Costa, “Adapting and Using Scrum in a Software Research and Development Laboratory” *Salesian Journal on Information Systems*, vol. 9, 2012, pp. 16–23.
- [7] T. Dingsøyr, T. Dybå and P. Abrahamsson, “A Preliminary Roadmap for Empirical Research on Agile Software Development” *AGILE 2008*, Toronto, Canada, 2008. doi: 10.1109/Agile.2008.50
- [8] K. Graßer and R. Freisler, “Agil und erfolgreich führen” *managerSeminare Verlags GmbH*, 2017.
- [9] A. Przybyłek, M. Olszewski, “Adopting collaborative games into Open Kanban” *Federated Conference on Computer Science and Information Systems (FedCSIS'16)*, Gdansk, Poland, 2016. doi: 10.15439/2016F509
- [10] A. Przybyłek, D. Kotecka, “Making agile retrospectives more awesome” *Federated Conference on Computer Science and Information Systems (FedCSIS'17)*, Prague, Czech Republic, 2017. doi: 10.15439/2017F423
- [11] A. Przybyłek, M. Zakrzewski, “Adopting Collaborative Games into Agile Requirements Engineering” *13th International Conference on Evaluation of Novel Approaches to Software Engineering (ENASE'18)*, Funchal, Madeira, Portugal, 2018. doi: 10.5220/0006681900540064
- [12] Website of txtUML project, <http://txtuml.inf.elte.hu/wiki/doku.php>
- [13] Website of Slack communication tool, <https://slack.com/intl/en-hu/>
- [14] txtUML project on gitHub, <https://github.com/ELTE-Soft/txtUML/>
- [15] Quora, <https://www.quora.com/Do-you-know-of-research-teams-adapting-Agile-frameworks-scrum-kanban-XP-etc-and-or-Design-Thinking-techniques-for-managing-research-projects>
- [16] ProjectManagement.com: [Applying-Scrum-to-Research-Projects](https://www.projectmanagement.com/blog-post/43810/Applying-Scrum-to-Research-Projects), <https://www.projectmanagement.com/blog-post/43810/Applying-Scrum-to-Research-Projects>
- [17] ProjectManagement.com: [What-should-Scrum-look-like-in-2019-](https://www.projectmanagement.com/blog-post/50228/What-should-Scrum-look-like-in-2019-), <https://www.projectmanagement.com/blog-post/50228/What-should-Scrum-look-like-in-2019->



## Real-Life Challenges in Automotive Release Planning

Kristina Marner  
Dr. Ing. h.c. F. Porsche AG,  
Porschestraße 911,  
71287 Weissach, Germany  
Email: kristina.marner@porsche.de

Sven Theobald  
Fraunhofer IESE,  
Fraunhofer-Platz 1,  
67663 Kaiserslautern, Germany  
Email:  
sven.theobald@iese.fraunhofer.de

Stefan Wagner  
University of Stuttgart,  
Universitätsstraße 38,  
70569 Stuttgart, Germany  
Email: stefan.wagner@iste.uni-  
stuttgart.de

□ **Abstract—Context:** The use of agile software development is increasing, even in regulated domains like the automotive domain. At the same time, traditional sequential processes are still in use. Collaboration between agile and hybrid projects within these complex traditional product development processes is difficult. Especially the creation and synchronization of a qualification phase plan is challenging. **Objective:** The aim of this study is to provide insights into the state of the practice to understand challenges related to the combined use of agile and traditional paradigms in release planning in the automotive domain. **Method:** Based on semi-structured interviews, an online survey with 39 respondents was conducted at Dr. Ing. h. c. F. Porsche AG. **Results:** We present the challenges identified in release planning, such as lack of transparency regarding the status quo of related projects. Furthermore, we motivate how agile development methods could improve collaboration between projects in release planning. **Conclusions:** There are many challenges in the context of co-existing agile and traditional projects. We discuss how agile practices like daily standup or continuous integration could address the identified challenges.

### I. INTRODUCTION

**T**ODAY, it is a competitive advantage to develop and put products on the market as early as possible. Agile software development methods and practices are commonly used to achieve this goal [1]. Practitioners want to benefit from increased project visibility, faster response to change, and shorter time to market [1] by adopting agile development practices.

Nonetheless, traditional approaches like the waterfall or the V-model are still predominant in highly regulated domains. Within these domains, the adoption of agile practices is hard to achieve and even not always desired [2]. To overcome the factors that hinder an agile transformation, regulated domains prefer adopting single agile practices [3] into their development processes [4]. This inevitably leads to a mixture of different development processes ranging from completely traditional processes to agile adaptations [5][6], which in turn results in more and more complex interfaces

[7] between all involved methodologies. The mixture of traditional and agile development practices is called hybrid development approaches [5]. Such approaches are commonly used in the automotive domain [8].

In the automotive domain, the complexity of software and systems is constantly increasing [9]. As automotive projects are generally large projects with many subprojects and suppliers, it is necessary to preserve the benefits of the existing rich development processes [10] to coordinate all involved parties. In addition, current software development in the automotive domain is intended to address safety-critical functionality by means of standardized processes to satisfy requirements given by the law.

Thus, it is a challenge to speed up software release cycles [2]. Creating and updating a common release plan that considers all dependencies is challenging, even more so when multiple parties work with different processes.

The aim of this work is to investigate the challenges in the release planning of automotive projects when traditional and agile processes co-exist.

The contribution of our work is as follows: We identify and analyze challenges in the qualification phase to identify improvements in the context of co-existing agile and traditional projects from the perspective of an automotive Original Equipment Manufacturer (OEM).

The remainder of this paper is structured as follows: Related work is presented in Section 2. Section 3 defines the research approach including the research questions and design, the research site and the participants, the data collection and analysis procedure, as well as the data collection instrument. The survey results are reported in Section 4. We conclude our work and outline future research in Section 5.

### II. BACKGROUND AND RELATED WORK

In the automotive domain, a hybrid project environment consists of two conflicting parts. There is the strategic framework on one side consisting of processes with many milestones planned a long time in advance before projects related to production and distribution go live. This strategic framework represents the time and content requirements, such as the product development process and thus defines a

□ This work was not supported by any organization

superordinate process. On the other hand, there is the operational level, where projects are performed in the way that best fits the project's character. On this level, projects are developed in an agile, hybrid or traditional way. A solution has to be found that synchronizes both levels and which enables coordinated release planning.

The automotive domain is a strongly regulated domain. Therefore, this combination cannot start in a green field, as strategic frameworks define different phases of the development process.

The Qualification Phase (QP) is the repetitive integration and testing process of an Electronic Control Unit (ECU) network, its sensors and actuators.

This phase is typically defined at the beginning of a project. The maturity level is determined to release the ECU network for further testing, usage, and development. The maturity levels provide information about the development progress of functions and ECUs in relation to the target state.

The Additional Qualification Phase (AQP) is an extra qualification phase with a reduced testing scope if the level of maturity is found to be insufficient and refers to a reduced scope of ECUs. The reduced test scope refers to the inadequate target state and is defined application-specifically. An AQP is not planned in advance but established depending on the quality level of the QP. In such cases, it is necessary and has to be executed. The selection of the test cases and the duration of the tests strongly depend on the errors identified during the QP.

Release planning in a hybrid project environment has barely been considered to date in the literature. Software release planning matches features to releases under the condition that different types of constraints are considered [11]. Heikkilä et al. [12] identify "an obvious gap in the research of release planning in large-scale agile software development organizations" in a literature review. However, they did not consider the combination of releases consisting of software and hardware.

Sax et al. [13] describe software release and configuration management in the automotive domain. Bestfleisch, Herbst and Reichert [14] define requirements for controlling and monitoring dependencies on other release processes with the help of workflow support. Müller et al. [15] define requirements for IT support to improve release management in the automotive domain. Lindgren et al. [16] identified key aspects of release planning in the context of software and system development projects. Furthermore, they captured the state of the practice for release planning in industry.

There is literature dealing with release planning in agile software development projects, both for single projects and for scaled projects. Danesh et al. [17] evaluated the methods used by companies to plan new software releases. Heikkilä et al. [18] present a case study where the agile release planning process in a scaled Scrum environment was evaluated. Heikkilä et al. [19] describe the qualification phase and present a case study of multi-team agile release planning with the help of this practice.

Karvonen et al. [20] conducted a systematic literature study to identify agile release engineering practices. Ameller

et al. [21] conducted a literature study to report on software release planning models. Overall, there is no direct related work that considers release planning in co-existing traditional and agile processes in the automotive domain. Some work deals with agile release planning, but none of the identified sources deals with the targeted hybrid project environment.

The HELENA study [5][6][8] investigates the combined use of agile and traditional practices in hybrid processes, but does not consider the co-existence of agile and hybrid projects and their synchronization. Theobald and Diebold [7] investigate and classify problems at the interface of agile development and a traditional environment. The work of this paper can be classified in the problem field "project planning" at the interface "project team" [7].

The focus of the majority of publications on release planning models are various kinds of mathematical models and simulations [22], which are ineffectual in complex industries [23]. Practitioners reported that these approaches are either too simple to generate a benefit or so difficult that they cannot reconstruct the whole process created [24][25].

There is a research gap considering hybrid project environments where projects with different development paradigms meet. Our paper aims to address this research gap.

### III. RESEARCH APPROACH

#### A. RESEARCH QUESTIONS

This paper aims to answer the following research question: *What are the challenges and consequences of the qualification phase in an automotive hybrid project environment?* To answer this question, three research questions were defined:

- RQ1. What are challenges concerning the qualification phase in a hybrid project environment?
- RQ2. What are the specific challenges of agile projects embedded in a traditional development context?
- RQ3. How could agility address the identified challenges?

#### B. RESEARCH DESIGN

To answer the research questions, we selected a two-step research approach. First, we set up an exploratory, qualitative interview study within a German automotive OEM. An interview guide for identifying challenges and problems with regard to the release planning process was specified. The interview guide was tested in a pilot interview. Emerging issues, such as vague phrases, were addressed before the qualitative interview study was conducted. In the second step, an online survey questionnaire was developed to validate the challenges identified from the qualitative interview study in detail.

The data collection instrument was a questionnaire containing 31 questions. The survey questionnaire contained open and closed questions structured into six categories (cf. **Table 1**).

TABLE 1. SURVEY QUESTIONNAIRE

Category	ID	Question
Context	1	What is your current role? [free text ]
	2	How long have you been working in that role? [free text ]
	3	What are you working on in your project? [E/E ECU, software component, function, connect service, vehicle project]
	4	Please select a sector to classify your project. [powertrain electronics, body electronics, infotainment, project is safety-critical, others]
	5	What kind of development method do you use? (agile, hybrid, or traditional) [use of adapted agile methods, hybrid methods, traditional approaches]
	6	If you are using agile or hybrid methods, please specify the method. [free text]
Qualification Phase	7	What do you think about the current number of qualification phases (incl. additional qualification phase)? [too high, adequate too low]
	8	How often are you able to generate current software versions ready to deliver? [never, seldom, often, always]
	9	Do you receive feedback about the qualification phase on time? [never, seldom, often, always]
	10	How often should a qualification phase take place in order for you to be ready to deliver? [every week, once a month, every 3 months, at larger intervals]
	11	Would additional releases in terms of partial composites with reduced test scope be helpful for safeguarding dependent ECUs? [yes, partially, no]
Planning	12	Is an initial planning of content possible? [never, seldom, often, always]
	13	Does an initial planning of content make sense? [never, seldom, often, always]
	14	How often is the content of the initial planning still up-to-date at the beginning of a qualification phase? [never, seldom, often, always]
	15	How difficult is it to get planning information for the relevant counterparts? [very difficult, difficult, easy, very easy]
	16	To what extent do management decisions, external influencing factors, or externally determined decisions influence your development process? [no impact, weak impact, strong impact, very strong impact]
Integration	17	To what extent does bug fixing affect the timely implementation of planned functionalities for the next qualification phase? [no impact, weak impact, strong impact, very strong impact]
	18	It is inevitable that software versions are released that are suboptimal concerning quality or content. [yes, partially, no]
	19	What kind of activities dominate your daily routine during a qualification phase? [free text]
	20	Rate the following statement: Additional qualification phases are necessary. [yes, partially, no]
	21	Rate the following statement: Additional qualification phases are reasonable. [yes, partially, no]
Coordination	22	Is the status of development transparent to you at any time? [yes, partially, no]
	23	Is the status of development of your stakeholders transparent to you at any time? [yes, partially, no]
	24	How important is the transparency of the development status of your relevant counterparts to you? [totally unimportant, unimportant, important, very important]
	25	Rate the following statements: - Stakeholder/Interfaces are known [Disagree, rather disagree, rather agree, agree] - Quality of coordination is good. [Disagree, rather disagree, rather agree, agree]
Testing	26	Development can no longer handle the high number of bug reports. [Disagree, rather disagree, rather agree, agree]
	27	Problem resolution management can no longer handle the high number of bug reports. [Disagree, rather disagree, rather agree, agree]
	28	What are the reasons for the high number of tickets? [free text]
	29	Do all planned changes to the ECU network have to be fully tested for each qualification phase? [yes, partially, no]
	30	Do all types of tests have to be performed for every ECU for each qualification phase? [yes, partially, no]
	31	When do all ECUs have to be fully tested? [every qualification phase, depending on the changes, not mandatory]

The categories and questions were derived from the insights gained in the previous interviews. The questions were originally written in German. The questionnaire went through four review cycles by an independent researcher as well as by a specialist from the case company. Review comments were discussed by the authors and addressed to improve the questionnaire.

In the first category, we elicited the “Context”, such as role and experience of the participant, as well as project type, area, and the development method used (traditional vs. agile). The second category, “Qualification Phase”, aimed at evaluating how many qualification phases are feasible. The third category, “Planning”, was for evaluating the need to have an initial plan as well as external influences on such a plan. At a certain point in the development process, an initial planning of the functional scope of an ECU must be submitted for each release. In addition to general ECU information, deviations from the required functional, network and diagnostic maturity levels must also be specified. We examined the need for AQPs in the fourth category “Integration”. Integration is an upstream part of the actual process and represents the integration of one or more ECUs into a whole network. Transparency of the status quo and the quality of coordination were the focus of the fifth category, “Coordination”. Finally, we covered all questions related to “Testing” in the last category, trying to evaluate which kind and intensity of tests are necessary and if and why there are so many bug reports. The test phase focuses on the execution of the qualification phase and is therefore a main activity.

#### C. DATA COLLECTION AND ANALYSIS PROCEDURE

To identify the main challenges, the first researcher conducted 26 semi-structured interviews, which took between 30 and 60 minutes each. The information from each interview was incorporated into later interviews. Because these interviews did not allow for quantitative results, an online survey was conducted to confirm the challenges and to draw a more complete picture by consulting different participants. This allows for quantitative results, but gave every participant the chance to provide further qualitative results by sharing their experiences.

95 potential participants were selected based on their roles, to cover all perspectives. Then the participants were invited via an email motivating the goal of the study and outlining the contents and the time expected to answer the questionnaire. A reminder email was sent after one week. Also, one of the participants forwarded the questionnaire to an additional group of 25 people. The survey was open from November to December 2018.

After extracting the data from the online survey tool<sup>1</sup> into an Excel document, we analyzed the answers for completeness. There were 39 complete responses, meaning all six pages of the questionnaire had been answered and thus the survey had been officially finished. In addition,

there were 16 incomplete answers where the questionnaire was not finished. Of these 16 incomplete answers, 1 participant stopped after category 3 (Planning), another one stopped after category 5 (Integration), and all others had discontinued the questionnaire even earlier. Although we had access to the incomplete data sets, we decided to only consider the complete data sets for further analysis. Since the survey was distributed to 120 people with 39 respondents, our response rate was about 33%. Afterwards, we conducted a descriptive analysis of the individual questions and analyzed the textual answers to identify common opinions.

#### D. RESEARCH SITE AND PARTICIPANTS

This study was conducted at Dr. Ing. h. c. F. Porsche AG, a manufacturer that builds sports cars for everyday driving. The division EE within Dr. Ing. h. c. F. Porsche AG in Weissach, Germany, is responsible for the development process of electronic systems and its integration into the development process of the complete vehicle. For achieving this goal, transparent development, processes and hence accurate release planning are essential.

The target population of our survey included all roles involved in the qualification phase process of automotive products where the subprojects differed in terms of the development approaches used, including agile as well as traditional methods. The sample selected consisted of stakeholders from Dr. Ing. h. c. F. Porsche AG involved in release planning activities. The participants were expected to be motivated enough to answer the comprehensive questionnaire because they anticipated improvements based on the findings that reflect their current situation.

#### E. THREATS TO VALIDITY

As the results only represent one specific case, it might not be possible to generalize them. However, the fact that the case company has the same framework conditions (regulated domains, complex supplier relationships and high safety requirements) as similar OEMs, others could benefit from the findings. The issues that were identified in the earlier interviews were addressed in the questionnaire, whereas new survey participants did not have a chance to add more individual problems during the online survey. There might be a bias concerning the stakeholders who participated. Some roles are overrepresented, while other relevant roles were not represented by many participants. This might have led to results that are skewed towards the opinion of certain roles. Nonetheless, many different roles participated in the study, providing answers from many perspectives. As in all surveys, non-response bias could have led to missing the opinions of certain participants.

#### IV. SURVEY RESULTS

This section contains the demographics and context of the respondents, followed by the presentation and discussion of the results of this work structured along the research questions.

<sup>1</sup>[www.limesurvey.org](http://www.limesurvey.org)

## A. CONTEXT

The respondents' professional experience in their current role (Q1) was slightly below six years on average, with a minimum of one year and a maximum of 16 years (Q2). Most of the respondents had management roles (n=17; 44%), others were responsible for projects, products, functions, integration, testing, quality, data, processes, or other related disciplines. 10 participants (26%) represented the operational level. The remaining 12 respondents (30%) had roles with responsibilities related to the environment of qualification phases.

The respondents described their working environment using one or more categories (Q3). Most participants reported working in vehicle projects (n=24), development of E/E components (n=18), development of functions (n=14), development of software components (n=12), and connected services (n=8). Others (n=7) dealt with IT backend, cross-project integration, distributed functions, or quality.

14% of the respondents answered that their project was safety-critical. Most participants assigned their project to the area of infotainment (n=13), followed by electronics for car bodies (n=11) and electronics for engines (n=7). Regarding the 24 additional classifications, ten participants reported working on crosscutting topics (Q4).

Most respondents reported using traditional development or project management approaches such as the V-model or sequential approaches (n=26). Only six respondents used adapted agile methods, and seven persons used hybrid approaches, which was defined as strongly adapted agile methods or use of only single agile practices (Q5). This showed that only one third of the study participants were using agile concepts at the time.

Agile implementations were based on Scrum or the Porsche-specific adaption of agile methods. One person even reported scaled agile and lean at the unit level combined with an adapted Scaled Agile Framework (SAFe). Single agile practices like daily standups, user stories, backlogs, retrospectives, or the Scrum Master role were used in traditional projects. Some respondents reported using both agile and traditional approaches at different project levels. One answer stated that agile was being used at the team level together with the V-model for whole projects, while another respondent reported using a sprint-like approach within the V-model due to highly dynamic changes in requirements. Another respondent indicated the use of different development paradigms in different life cycle phases (Q6).

## B. RQ1: CURRENT CHALLENGES

In the following, the current challenges will be presented and discussed along the categories of the survey questionnaire. RQ1. What are challenges concerning the qualification phase in a hybrid project environment?

### 1) QUALIFICATION PHASE

The majority of the participants (n=22; 56%) stated that the current number of releases (p.a.), including all additional

qualification phases and special qualification, is too high (Q7). On closer inspection, there is a discrepancy between the answers by managers and those by developers with responsibility for products or functions. The former (n=17) reported that the existing number of releases is too high (56%), while the latter said it is too low (25%).

An analysis of the comments field of this question shows results relating to the regulated defined number of releases. The developers confirmed their opinion and asked for a higher number of qualification phases. The management group agreed with the regulated defined numbers.

Further information concerning the ordinary number of qualification phases was given by the group of developers using agile methods. For the majority of those participants, the absolute number of qualification phases is too low to use agile methods properly.

The next issue concerned the delivery results in the required form (Q8). 60% of the survey participants answered that the required deliverable is seldom available in the required quality. In contrast, 40% replied that it is always or at least most of the time possible to create a delivery version for every requested release.

74% of the participants answered that they mostly receive feedback about qualification phases on time (Q9). The next question dealt with the number of qualification phases with regard to generate software version (Q10). Two-thirds of the participants stated that qualification phases should take place at least each quarter of the year. In contrast to the last question (Q11) in this category, 46% called for additional qualification phases with reduced test scopes.

## 2) PLANNING

This category highlights the characteristics around planning. The first question (Q12) aimed at evaluating the feasibility of initial planning at the beginning of the project. 50% of the participants in our study reported that initial planning is possible, and the other half answered that such a plan is rarely possible. At the beginning of a project, the decisions for or against a supplier have sometimes not been made yet. That is one reason why it is difficult to generate an initial planning. Another person replied that requirements for functions are the results of testing, which is done further on in the development process.

In a further question, the participants were asked if such initial planning would be meaningful (Q13). A significant majority (74%) stated that planning at the beginning of a project is reasonable because it is a resilient starting point for further steps. Participants also mentioned the existing change management process, which permits updates at any time.

The next question (Q14) regarding this topic dealt with the projected content before the next release in terms of timeliness. The results show that scheduled content is frequently impossible to implement in practice (80%). The majority of the participants stated that awareness still exists for high quality in planning. Planning updates have to pass a committee, which is one reason why change requests are not implemented in the current release. Also some areas,

“connected car”, are very dynamic, which is another reason for the bad current state of planning, which is not up-to-date. Receiving information about planning details from the relevant stakeholders is perceived as challenging (Q15). 74% of the respondents replied that obtaining information on time is difficult because there are no regulated tasks nor a consistent workflow for changing the relevant information.

Another issue is the impact of management decisions during the development cycle (Q16), which implies that these cannot be implemented easily. 90% of the respondents rated this influence as strong or very strong and reported that the development of new functionality suffers from having to deal with unexpected changes demanded by management. Some respondents complained about management decisions that change the backlog priority and have severe effects on further procedures.

### 3) INTEGRATION

This category contains the results relating to the challenges of software and hardware integration during a development cycle.

During a qualification phase, new software versions are tested at different levels of integration. The test results and even bug fixing have a great impact on the subsequent procedure (Q17). 87% of the participants answered that bug fixing affects their timely implementation related to the next release. Because there is no hold available in the project plan, this even leads to delays of the next scheduled functions (Q18).

Another question in this category dealt with the activities during a qualification phase (Q19). The main activities or tasks linked to the respective role are: Management is engaged in coordination and ensuring the scheduled scope with regard to the next release. At the operational level, tracking of test results and analysis of upcoming bug tickets are the main concerns. Both groups have to handle the subsequent deliveries.

Almost all interviewees (96%) admitted that delivering software versions with high quality is infeasible when they also have to provide the content planned for the next release. The results considered for integration have low maturity, due to the increasing pressure of costs and deadlines.

For this reason, additional qualification phases have been established subsequent to the original deadline. We wanted to know if such additional qualification phases are necessary (Q20) and reasonable (Q21). 65% of the participants considered additional qualification phases necessary and 35% were convinced that they are reasonable.

The main reasons given by the participants for subsequent integration were poor software quality, lack of adherence to delivery dates on the part of the suppliers, poor scheduling without buffering, and no complete bug fixing from the previous qualification phase.

### 4) COORDINATION

Transparency and coordination were the relevant aspects in this category (Q22). We asked whether the current

development status of the respondents' own team or dependent teams is sufficiently transparent. Only 26% (n=9) reported that their own development is transparent. The majority of the respondents rated transparency as only partial (n=20; 51%) or non-existent (n=10; 26%).

Next, the results of questions Q23 and Q24 are presented. The questions dealt with the transparency of the status of projects by relevant stakeholders and relevant counterparts. Here, only 15% (n=6) of the respondents answered that the development status of other projects is transparent for them. Most participants (n=19; 49%) reported partial transparency, while 36% (n=14) reported a lack of transparency. Reasons for the lack of transparency were missing time and coordination mechanisms, and the use of outdated content of the release plans.

The transparency of the status quo of a certain development project is very important and closely linked to the quality of a release. 95% of the respondents supported the statement that having a transparent software version at any time is important. It is necessary due to the complexity, dependency, and connectivity of software engineering.

Another question aimed at getting information about the communication structures within the company and involved persons from the release planning process (Q25). The participants had to rate whether they knew their interfaces and relevant stakeholders and whether the quality of the coordination was good. This rating had to be done for several interfaces: within the team, between team and testing, within the case company, within the company group, as well as towards external suppliers.

The results presented in Fig. 1 (bottom figure) demonstrate that communication quality decreases with longer communication paths: Communication within a project was perceived as good, but the quality was perceived as decreasing in communication within the company and even worse in communication with suppliers (internal means company group and external suppliers). Similarly, the relevant stakeholders and interfaces of the wider project context were reported less known than those within the team (see Fig. 1, top figure).

### 5) TESTING

This category assesses the testing situation. The first question aimed to evaluate whether the number of bug reports is still controllable by development (Q26) or problem resolution management (Q27). Overall, 56% (n=22) of the participants agreed (n=7) or rather agreed (n=15) that development is able to control the high number of bug reports. The remaining respondents had a tendency to disagree (n=9) or disagreed (n=8).

Concerning problem resolution management, most participants (n=25) disagreed (n=8; 21%) or had a tendency to disagree (n=17; 44%). The minority of the participants agreed (n=5; 12%) or rather agreed (n=9; 23%)

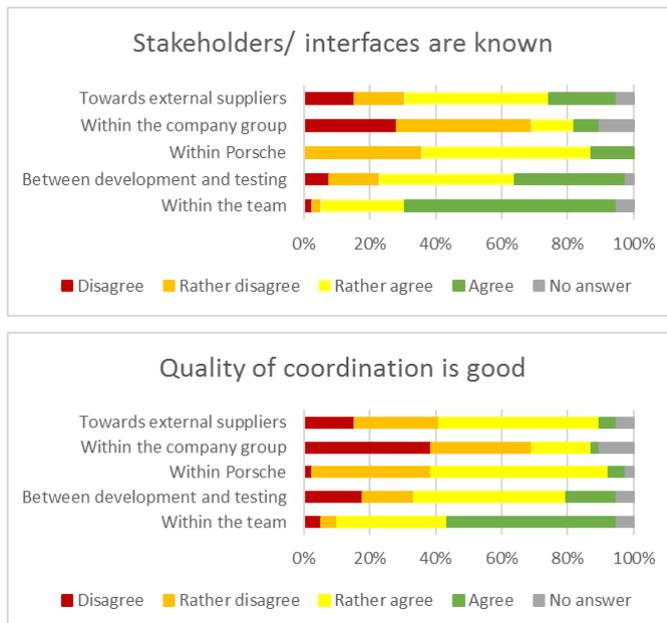


Fig. 1. Known interfaces (top) and quality of coordination (bottom)

Furthermore, the participants were asked about reasons for the high number of bug reports / tickets (Q28). The survey revealed that identifying errors is usually not done before the upcoming release due to insufficient development time, cost, and deadline pressure. It was reported that the intensity of testing by the supplier was not adequate. Other reasons given for the high number of error tickets are the rising complexity of the product itself, the lack of coordination within the team, and inadequate requirements engineering. Generally, it can be stated that the software quality before a qualification phase is insufficient and questionable, endangering the success of the qualification phase.

Software changes may have severe effects on interfaces, which is why tests have to be done. The need for testing the software changes to the full extent for every qualification phase (Q29) was not seen by 18% of the respondents (n=7), who claimed that this is not necessary. Most respondents (n=18; 46%) said that changes have to be tested to the full extent for every planned release. The remaining 36% (n=14) partially agreed that testing is always necessary and specified in the comments specific situations where more testing was necessary or less testing was acceptable. Some stated that the scope of testing depends on the number of changes made or on the development phase. One respondent commented that it is not possible to test all changes; another one said that full testing is always necessary because cross-dependencies only become visible by testing within a release.

Only 10% (n=4) of the respondents agreed that all types of tests have to be performed in every release cycle (Q30). 39% (n=15) disagreed with this statement and about half of them (n=20; 51%) partially agreed. The participants further pointed out that conducting all tests is not feasible or that the necessary types of tests are predefined in the test strategy and depend on the change itself. Others reported that regression tests are often sufficient, or that full releases need to be tested more accurately than partial releases.

To save testing effort, it is important to know when comprehensive testing (including all types of tests) of all ECUs needs to be done (Q31). 85% of the respondents (n=33) answered that testing needs to be done dependent on the software, hardware, or functional changes. Five respondents (13%) claimed that the ECUs have to be tested once per qualification phase, either at the beginning or at the end. 2% (n=1) said that testing is not always necessary. One participant commented that due to the high product complexity and low software quality, all ECUs have to be tested as an integrated system with all possible tests, or at least with good regression tests. Another one claimed that comprehensive testing is not possible for all system parts, but major parts can be covered with a good testing strategy.

### C.RQ2: AGILE-SPECIFIC CHALLENGES

Existing vehicle development processes emerged at a time when agility was not present yet and were formalized based on traditional development methodologies. Due to the regulations, strict production deadlines and the complexity in vehicle development, the need to have formal processes will remain. However, the potential to integrate agile processes must be evaluated in order to exploit the benefits of agility. New technologies such as cloud services implicate a stronger customer focus, to be able to respond more flexibly to customer needs, which results in conflicts with the slow and unresponsive traditional development. Innovation is happening fast in the automotive company, and companies have to react in time to stay competitive.

Iterative cycles are already incorporated into many processes, but other concepts of agile methods initially designed for small teams are more difficult to integrate or synchronize with the existing rigid processes. The OEMs are currently performing a balancing act between fixed framework conditions and scope for flexibility. On the one hand, legal requirements, standards and production requirements must be observed and on the other hand, developers want to act more freely without being restricted by guidelines. The results of this survey indicate that this is not a simple procedure.

The survey revealed that if departments are already working with agile methods, they only use them to a certain extent. Our initial expectation was that agile methods are commonly used at least in fields such as connected car, with its digital services and shorter development cycles. The differences between our expectations and reality may be caused by the lack of a common understanding of agile methods. This is confirmed by the inconsistency of the answers by the respondents, who considered additional qualification phases necessary but at the same time did not demand more qualification phases. The reason for this may be a lack of knowledge about agile methods.

There is also a lack of suitable means of communication for short, regular exchanges aimed at establishing transparency between all participants. Such possibilities for fast feedback would also increase the overall quality of voting and benefit the flow of information. Respondents

(n=7) from agile projects reported that the length of release cycles is too long and does not suit agile approaches.

The fact that management decisions have such a strong influence on the further course of development illustrates that decisions are made at higher levels of hierarchy. In an optimal agile environment, the development team makes the decisions. Based on the priorities set by the Product Owner and the requirements dependencies identified by the development team, a Scrum team knows best how to achieve the best solutions. At the beginning of each iteration, they commit to a product increment that is valuable and achievable. If management forces decisions upon the team during an iteration, results can be expected to be suboptimal.

However, this is only the point of view of a single team. If each single team cannot meet their commitments, the qualification phase of an integrated product is going to raise problems. One reason is that the release plan, which considers dependencies between different projects, gets unofficially changed without being updated. That means the developers change their release plans on the operational level without having the change approved and without informing the affected interfaces.

#### D.RQ3. IMPROVEMENTS WITH AGILE METHODS

There are many challenges that are predestinated to be solved with agility. The survey revealed that transparency and coordination are highly important for a successful qualification phase. Some of the interviewees states that the communication path in their department is too long, which causes loss of time and a lack of coordination. This argument is supported by the fact that some of the participants reported not knowing their interfaces and relevant stakeholders, resulting in bad synchronization and integration structures. By using agile development and small working groups with no typical hierarchy, interface management and short communication paths could become possible [26].

Currently, additional qualification phases are started to fix the remaining bugs or to finish some functionalities that had been planned for the previous release cycle. Due to the increased effort for these activities, the planned results for the next release cycle cannot be fully achieved, pushing a wave of additional efforts, e.g., for coordinating additional qualification phases, through the whole project. Increased transparency regarding the content that was finished in an iteration can be achieved with a definition of done and by incorporating time-boxed sprints. At the end of each sprint, the status quo is assessed, and unfinished requirements can be planned for the next sprint.

Another characteristic of sprints is that requirements are usually not changed, especially not from outside the team. This would also help to stabilize the release plan, which would help to achieve higher-quality products delivered for integration by each single team. Sprints are usually short iterations of several weeks. Respondents from agile projects reported that the length of release cycles is too long, and that they would prefer receiving feedback earlier. This issue leads to work overload and defined timelines not being

achieved, which ultimately leads to lower software quality. In addition, development costs increase due to many additional qualification phases. By using agile methods and more intermediate steps, including regular assessments of the project state, discrepancies could be identified earlier.

Agile teams use face-to-face communication and daily standups to synchronize their work in order to achieve their sprint goal. In a scaled environment, so-called Scrum-of-Scrums are scaled daily standups where representatives of different teams synchronize their development status and plan their dependencies. The Scaled Agile Framework (SAFe) uses an architectural runway to coordinate architectural decisions between the single development teams to facilitate integration.

Continuous integration is commonly used in agile projects and could be of benefit in qualification phases. Integrating smaller work products incrementally can replace a larger and more complex final integration and provides early transparency about the finished content of the release as well as raising awareness of dependencies.

In general, regular retrospectives can be held at the end of each sprint, helping the team to raise issues impeding their work and improve their development process. Conducting retrospectives together with relevant stakeholders and dependent projects helps to continuously improve collaboration between teams.

#### V. CONCLUSION AND FUTURE WORK

Agile development is being increasingly adopted even in regulated domains such as the automotive domain. There it has to be synchronized with co-existing traditional development approaches. Qualification phases of traditional, hybrid, and agile projects are difficult. An online survey was used to identify challenges in the release planning of a German automotive OEM. The results show that communication and information issues such as inadequate familiar counterparts predominate in the case company. Dependencies between input and output relations are key results, too. Furthermore dissatisfied software quality during the system tests leads to further challenges. Another key statement of the survey results is the limited capacity to act to the supplier relationships.

We presented the main challenges in detail and discussed the state of agility and the conflicts arising in the context of co-existing traditional and agile approaches. We outlined how agile concepts could improve some of the identified challenges and thus provided recommendations for practitioners.

In the future, we plan an in-depth analysis and interpretation of the results, including a more detailed analysis of the questions, by trying to identify further correlations. With a questionnaire adapted to collect experiences outside the case company, we want to check whether there are similar problems at other automotive companies or even companies from other regulated domains that are developing complex systems in a hybrid project environment.

## REFERENCES

- [1] VersionOne: The 13th annual state of agile report. (2019). [www.col-lab.net](http://www.col-lab.net)
- [2] Hohl, P., Münch, J., Schneider, K., Stupperich, M. (2016). Forces that prevent agile adoption in the automotive domain. In *proc. of the International Conference on Product-Focused Software Process Improvement, PROFES 2016: Product-Focused Software Process Improvement* (pp 468-476). DOI.org/10.1007/978-3-319-49094-6\_32.
- [3] Diebold, P., Zehler, T. (2016). The right degree of agility in rich processes. In *Managing Software Process Evolution* (pp. 15-37). DOI: 10.1007/978-3-319-31545-4\_2.
- [4] Diebold, P., Theobald, S. (2018). How is agile development currently being used in regulated embedded domains? In *proc. of JSEP'18*. DOI.org/10.1002/smr.1935.
- [5] Klünder, J., Hohl, P., Fazal-Baqaie, M., Krusche, S., Küpper, S., Linssen, O., & Prause, C.R. (2017). HELENA study: reasons for combining agile and traditional software development approaches in german companies. In *proc. of PROFES'17*. DOI.org/10.1007/978-3-319-69926-4\_32.
- [6] Kuhrmann, M., Diebold, P., Münch, J., Tell, P., Garousi, V., Felderer, M., Trektere, K., McCaffery, F., Linssen, O., Hanser, E., Prause, C. R. (2017). Hybrid software and system development in practice: waterfall, scrum, and beyond. In *proc. of the 2017 International Conference on Software and System Process* (pp. 30-39). <https://doi.org/10.1145/3084100.3084104>.
- [7] Theobald S., Diebold P. (2018): Interface Problems of Agile in a Non-agile Environment. In: Garbajosa J., Wang X., Aguiar A. (eds) *Agile Processes in Software Engineering and Extreme Programming, XP 2018. Lecture Notes in Business Information Processing*, vol 314. Springer, Cham. DOI.org/10.1007/978-3-319-91602-6\_8.
- [8] Tell, P., Klünder, J., Küpper, S., Raffo, D., MacDonell, S. G., Münch, J., Pfahl, D., Linssen, O., Kuhrmann, M. (2019). What are Hybrid Development Methods Made Of? An Evidence-based Characterization. In *proc. of the ICSSP19*. DOI.org/10.1109/ICSSP.2019.00022.
- [9] Broy, M. (2006). Challenges in automotive software engineering. In *proc. of the 28th international conference on Software engineering* (pp. 33-42). ACM. DOI.org/10.1145/1134285.1134292
- [10] Hohl, P. (2019). An assessment model to foster the adoption of agile software product lines in the automotive domain. University of Hannover, Hannover, Germany. DOI.org/10.1109/ICE.2018.8436325.
- [11] Saliu, O., Ruhe, G. (2005). Supporting Software Release Planning Decisions for Evolving Systems. In *proc. of NASE SEW-29*, Washington DC. DOI: 10.1109/SEW.2005.42.
- [12] Heikkilä, V. T., Paasivaara, M., Rautiainen, K., Lassenius, C., Toivola, T., & Järvinen, J. (2015). Operational release planning in large-scale scrum with multiple stakeholders—A longitudinal case study at F-Secure Corporation. In *Information and Software Technology*, 57, (pp.116-140). DOI.org/10.1016/j.infsof.2014.09.005.
- [13] Sax, E., Reussner, R., Guissouma, H., Klare, H. (2017). A Survey on the state and future of automotive software release and configuration management. *Karlsruhe Reports in Informatics*, 11. DOI: 10.5445/IR/1000075673.
- [14] Bestfleisch, U., Herbst, J., & Reichert, M. (2005). Requirements for the workflow-based support of release management processes in the automotive sector. In *proc. of the 12th European Concurrent Engineering Conference ECEC'05*.
- [15] Müller, D., Herbst, J., Hammori, M., Reichert, M. (2006). IT support for release management processes in the automotive industry. In *proc. of the International Conference on Business Process Management* (pp. 368-377).
- [16] Lindgren, M., Land, R., Norström, C., Wall, A. (2008). Key aspects of software release planning in industry. In *proc. of the 19th Australian Conference on Software Engineering* (pp. 320-329). DOI: 10.1109/ASWEC.2008.4483220.
- [17] Danesh, A. S., Ahmad, R. B., Saybani, M. R., & Tahir, A. (2012). Companies approaches in software release planning-based on multiple case studies. *JSW*, 7(2), 471-478. DOI:10.4304/jsw.7.2.471-478.
- [18] Heikkilä, V. T., Paasivaara, M., Lassenius, C., & Engblom, C. (2013). Continuous release planning in a large-scale scrum development organization at ericsson. In *proc. of the International Conference on Agile Software Development* (pp. 195-209). DOI.org/10.1007/978-3-642-38314-4\_14.
- [19] Heikkilä, V., Rautiainen, K., & Jansen, S. (2010). A revelatory case study on scaling agile release planning. In *proc. of the 2010 36<sup>th</sup> EUROMICRO Conference on Software Engineering and Advanced Applications* (pp. 289-296). DOI: 10.1109/SEAA.2010.37.
- [20] Karvonen, T., Behutiye, W., Oivo, M., & Kuvaja, P. (2017). Systematic literature review on the impacts of agile release engineering practices. *Information and Software Technology*, 86, (pp. 87-100). <https://doi.org/10.1016/j.infsof.2017.01.009>.
- [21] Ameller, D., Farré, C., Franch, X., & Rufian, G. (2016). A survey on software release planning models. In *International Conference on Product-Focused Software Process Improvement* (pp. 48-65). DOI: 10.1007/978-3-319-49094-6\_4.
- [22] M. Svahnberg, T. Gorschek, R. Feldt, R. Torkar, S. B. Saleem, and M. U. Shafique (2010). A systematic review on strategic release planning models. *Information and Software Technology* (pp. 237–248). DOI: 10.1016/j.infsof.2009.11.006
- [23] Carlshamre, P. (2002). Release planning in market-driven software-product development: Provoking an understanding. *Requir. Eng.* 7(3), (pp. 139–151). DOI: 10.1007/s007660200010.
- [24] Jantunen, S., Lehtola, L., Gause, D. C., Dumdum, U. R., Barnes, R. J. (2011). The challenge of release planning. In *proc. of the Fifth International Workshop on Software Product Management*, (pp. 36–45). DOI: 10.1109/IWSPM.2011.6046202.
- [25] Benestad, H.C., Hannay, J.E (2011). A comparison of model-based and judgment-based release planning in incremental software projects. In *proc. of the 33rd International Conference on Software Engineering*, (pp. 766–775). DOI: 10.1145/1985793.1985901.
- [26] Spiegler S.V., Heinecke C., Wagner S. (2019). Leadership Gap in Agile Teams: How Teams and Scrum Masters Mature. In: Kruchten P., Fraser S., Coallier F. (eds) *Agile Processes in Software Engineering and Extreme Programming, XP 2019. Lecture Notes in Business Information Processing*, vol 355. Springer, Cham. (pp 37-52) DOI: 10.1007/978-3-030-19034-7\_3



# On the Agile Mindset of an Effective Team – An Industrial Opinion Survey

Jakub Miler

Gdansk University of Technology  
Faculty of Electronics, Telecommunications  
and Informatics  
11/12 Narutowicza St., 80-233, Gdansk, Poland  
Email: jakub.miler@eti.pg.edu.pl

Paulina Gaida

Omida Finance Sp. z o.o.  
Grunwaldzka 472B St.,  
80-309 Gdańsk, Poland  
Email: paulina.gaida@gmail.com

□ **Abstract**— In this paper we present the results of an opinion survey among 52 agile practitioners who evaluated the importance of 26 selected elements of the agile mindset to the effectiveness of an agile team. In total, we have identified 70 unique agile mindset elements based on 11 literature sources and 5 interviews with industry experts. 7 elements belonged to the “support for business goals” category, 20 to the “relationships within the team” category, 24 to the “individual features” category, and 19 to the “organization of work” category. Our survey shows the relative importance of the selected 26 agile mindset elements according to our respondents which is not fully consistent with the principles behind the Agile Manifesto.

## I. INTRODUCTION

Agile Manifesto [1] together with the principles behind the Agile Manifesto [2] founded a set of driving values and key principles for the agile software development. Agile practitioners emphasize that effective performance of an agile team requires not only a given set of procedures, techniques and rituals, but, above all, a particular attitude, way of thinking and behavior of both the individuals and the entire team – a so called ‘agile mindset’ [3, 4].

Working in agile teams requires many non-technical and social competencies related to communication, organization, business, improvement and many more [5]. These are not the typical strong competencies among software engineers [6], which is why they require support of Scrum Masters, mentors and coaches to develop deep understanding of the fundamentals of Agile. Agile mindset, by addressing all of these competence areas and by suggesting important factors to the effective teamwork, supports practitioners in mastering Agile in their projects [4]. Altogether, developing the proper agile mindset contributes to the increasing success of agile software projects [7].

The principles behind the Agile Manifesto themselves [2] recommend such attitudes and behaviors as focus on customer satisfaction, openness to change, face-to-face communication, sustainable development, simplicity, self-organization and improvement by frequent reflection. The

□ This work was partially supported by the DS Funds of ETI Faculty, Gdansk University of Technology.

agile methods such as Scrum [8], Kanban [9], SAFe [10] and other elaborate these principles further on, however the evolution of the IT industry since the Agile Manifesto calls for deeper and more current insight into the concept of ‘being and working agile’. In our research, we assume the definition of ‘an agile mindset’ as a set of one’s attitudes, behaviors and ways of thinking that enhance their and their team’s effectiveness in working following the agile values and principles to the benefit of the customers.

This research aims at studying the elements of the agile mindset and their importance to the effectiveness of an agile team. We have formulated the following research questions: (RQ1) What agile mindset should the members of an agile team have? (RQ2) What is the importance of the particular agile mindset elements to the effectiveness of an agile team? (RQ3) What are the most important elements of the agile mindset to the effectiveness of an agile team?

The main contribution of this paper is the broad identification of the elements of agile mindset and the partial evaluation of their importance to the effectiveness of an agile team based on an industrial opinion survey. This extends the reviewed literature with deeper understanding of the concept of ‘agile mindset’ and the relative importance of its elements.

The paper is structured as follows. Section II presents our research method of identification, selection and evaluation of the agile mindset elements. Section III reports the results of the identification phase based on the literature review and the interviews with experts. Section IV presents the selection of the agile mindset elements for further evaluation. Section V reports the results of the survey together with the analysis of confounding variables and the comparison to the principles behind the Agile Manifesto. Section VI discusses the threats to the validity of this research, followed by the discussion in Section VII and conclusions in Section VIII.

## II. RESEARCH METHOD

Our research comprised three steps: (1) identification of the elements of an agile mindset and their categorization, (2) selection of the agile mindset elements for evaluation, (3) evaluation of the relative importance of the selected agile mindset elements to the effectiveness of an agile team.

The first step involved the review of current literature and the interviews with experts from industry. The literature review covered mainly grey literature (books, blogs, portals), as the scientific databases such as Scopus or Web of Science provided very few results. We have focused on Internet sources reporting on industrial practice or written by agile practitioners and published by renowned publishers or portals. In total, we analyzed 11 literature sources.

To identify the agile mindset elements more thoroughly, we have carried out 5 structured interviews with industry experts with 2 to 5 years of experience in agile teams. They mostly worked as developers and Scrum Masters with various agile methods. The characteristics of the interviewed experts are given in Table I. Experience is given in years.

TABLE I.  
CHARACTERISTICS OF THE INTERVIEWED EXPERTS

ID	Position	Exp.	Methods
A	developer	3	Scrum
B	developer, tester	2	Kanban
C	developer	2	Scrum
D	Scrum Master, Agile Coach	5	Scrum, Kanban, XP
E	Scrum Master	3	Scrum, Kanban, Scrumban

The interviews were carried out in late May – early June 2018 in a form of face to face meetings. Experts A to C were not provided the interview questions in advance, which resulted in limited answers. Thus, experts D and E were sent the questions before the interview, which allowed them to think over their answers and generally resulted in more original insight into the subject matter. We have followed the given interview guide:

I. Preliminary questions:

1. For how long have you been working in agile teams?
2. What methodology are you using in your projects (Scrum, Kanban, XP - Extreme Programming, others)?
3. What is your role in the team (developer, tester, Scrum Master, etc.)?

II. General questions about the philosophy of agility:

1. What is agility for you?
2. What does "agile mindset" mean for you?

III. Questions about agile mindset elements (at least 3 elements from each question):

1. Which beliefs do you think are necessary to have the agile mindset?
2. What are the most important values for a person with the agile mindset?
3. What principles should be followed by a person with the agile mindset?

IV. Questions about the importance of agile mindset elements (at least 5 elements from each question):

1. What are the most important attitudes, rules and behaviors at the interpersonal level in an agile team?

2. What are the most important attitudes, rules and behaviors in the work organization of an agile team?
3. What are the most important attitudes, rules and behaviors when dealing with customers in an agile team?

V. Questions about the impact of agile mindset on work efficiency:

1. What attitudes, behaviors and beliefs have the greatest impact on the efficiency of agile teams (name at least 5)?
2. Has your team worked inefficiently for reasons related to the agile mindset? What were these reasons?
3. Do you think it is necessary to have the agile mindset to work effectively in an agile team? Why?

Categorization of the identified agile mindset elements was done a posteriori based on keyword analysis in the results of the literature review. The same categorization was used for the interview results. The final list of identified agile mindset elements was elaborated by summing the sets of elements in the literature review results and interview results in each category, followed by merging the duplicates. We have noted the number of times each element was mentioned in the literature and the interviews (i.e. number of sources and number of experts, respectively, see Tables II and III).

The total number of identified agile mindset elements exceeded the capacity of a practical survey, so we had to select a subset of elements for further evaluation. As we aimed at one question per agile mindset element, we wanted to select no more than 30 agile mindset elements based on their frequency in sources (which is not importance). We have decided to include the elements found in at least 6 out of 11 literature sources or given by at least 2 out of 5 experts. These thresholds assume the majority of literature sources and some minimal agreement of the experts. Such thresholds favor the elements given by the experts, but this was our deliberate decision. Finally, such criteria resulted in 26 agile mindset elements selected for further evaluation. Other elements may be investigated in a separate study.

To evaluate the relative importance of the selected agile mindset elements to the effectiveness of an agile team, we have run a survey among agile practitioners in the IT industry. The survey was built on-line with Google Forms and distributed via e-mail, Facebook, forums etc. Respondents were asked to give their opinion on the degree to which a particular agile mindset element enhances the effectiveness of an agile team in the Likert-type 6 level scale of 0 to 5, where 0 meant "no impact" and 5 meant "key impact". The answers were optional which accounted for the cases of respondents' indecision or insufficient knowledge. The survey was organized by agile mindset categories. Additionally, we asked about the respondents' experience and their role in agile teams. Although basic Likert scale is ordinal, we used the Likert-type interval scale with assigned values of 0 to 5 in the survey and the data analysis [11].

III. IDENTIFICATION OF AGILE MINDSET ELEMENTS

A. Literature review

Using generic search engines such as Google, we have found the following literature on the topic of agile mindset:

1. “Agile Project Management: Managing for Success”, a book by James A. Crowder and Shelli Friess [12],
2. “The Agile Enterprise: Building and Running Agile Organizations”, a book by Mario E. Moreira [13],
3. “Being Agile: Your Roadmap to Successful Adoption of Agile”, another book by Mario. E. Moreira [14],
4. “The Agile Mindset – Making Agile Processes Work”, a book by Gil Broza [4],
5. “Five Agile Factors: Helping Self-management to Self-reflect”, a research paper by Christoph J. Stettina and Werner Heijstek [15],
6. “Learning Agile: Understanding Scrum, XP, Lean and Kanban”, a book by Andrew Stellman and Jennifer Greene [16],
7. “What Exactly is the Agile Mindset?”, an on-line article by Susan McIntosh for InfoQ portal [17],
8. “What does it mean to have an agile mindset?”, an on-line article by Leanne Howard for AgileConnection portal [18],
9. “It’s All About the Mindset”, an on-line article by Sayi Parvatam for Scrum Alliance portal [19],
10. “Fixed Mindset versus Agile Mindset”, an on-line article by V. Godugu for Scrum Alliance portal [20],
11. “Agile Is Not a Process, It’s a Mindset”, an on-line article by Lisa Rich for AgileConnection portal [21].

In total, we identified 58 elements of agile mindset in the literature. Table III lists these elements grouped into categories with the indication of relevant sources. The identifier of each element combines the “L” prefix (standing for the literature), the category symbol and the consecutive number of the element in each category. The list is ordered by descending number of sources in each category.

We have identified four categories of the agile mindset elements: (1) support for business goals, (2) relationships within the team, (3) individual features, (4) organization of work. The first category, denoted by G symbol in Table III, focuses on the product value and relations with the customer. The second category, denoted by the T symbol, covers the issues of collaboration and relations within the agile team. The third category, denoted by the I symbol, tackles the behavior and attitude of an individual in an agile team. Finally, the fourth category, denoted by the O symbol, involves the aspects of methods, techniques and rules.

B. Interviews with experts

The 5 interviews with experts A to E provided 16, 18, 16, 17, and 16 agile mindset elements, respectively. Repeating elements were merged. In total, we identified 39 unique agile mindset elements with the interviews. Table II lists these elements grouped into categories with the indication of

relevant sources. The identifier of each element combines the “E” prefix (standing for the experts), the category symbol and the consecutive number of the element in each category. The categories and their symbol are the same as in the literature review. The list is ordered by descending number of interviews in each category.

TABLE II. ELEMENTS OF THE AGILE MINDSET IDENTIFIED WITH THE INTERVIEWS

ID	Element name	Source	n <sub>E</sub>
E.G1	Cooperation with the customer based on partnership	B, C, D	3
E.G2	Attitude towards customer satisfaction and needs	B, D	2
E.G3	Continuous delivery of a valuable product in short intervals	A	1
E.G4	No assumption that the customer is always right	A	1
E.T1	Mutual trust	A, B, C, D, E	5
E.T2	Sincerity	A, B, C, E	4
E.T3	Helping each other	B, C, D, E	4
E.T4	Mutual listening	A, B, C	3
E.T5	Mutual respect	A, B, D	3
E.T6	Equality in the team	B, C, D	3
E.T7	Focus on achieving common goal	A, C, E	3
E.T8	Searching for a solution to the problem instead of finding the guilty	A, B	2
E.T9	Direct communication - face to face conversations	B, D	2
E.T10	Team responsibility	C, E	2
E.T11	Taking into account the opinions of other people	A	1
E.I1	Openness to change	A, B, C, D, E	5
E.I2	Positive attitude	A, B, E	3
E.I3	Continuous improvement and learning	B, C, E	3
E.I4	Being motivated	A, B	2
E.I5	Openness to criticism and feedback	A, D	2
E.I6	Openness to others	C, D	2
E.I7	Willingness to constantly acquire knowledge	B	1
E.I8	Pragmatism	B	1
E.I9	Individual initiative	B	1
E.I10	Courage	D	1
E.I11	Commitment	D	1
E.I12	Creativity, innovation	D	1
E.I13	Being a visionary	D	1
E.I14	Understanding the need for change	E	1
E.I15	Responsibility	E	1
E.I16	Understanding the significance of retrospectives	E	1
E.O1	Self-organization	A, C, D, E	4
E.O2	Finishing the current task before taking the next one	A, C, E	3
E.O3	Asking questions in case of insufficient knowledge	B, C, D	3
E.O4	Maintaining a steady pace of work	A, E	2
E.O5	Transparency in decision-making and actions	C, E	2
E.O6	Sharing knowledge and results	C	1
E.O7	Focus on the tasks performed	D	1
E.O8	Focus on cross-functional teams	E	1

TABLE III.  
ELEMENTS OF THE AGILE MINDSET IDENTIFIED IN THE LITERATURE

ID	Element name	Source	n <sub>i</sub>
L.G1	Continuous delivery of a valuable product in short intervals	[12], [13], [14], [4], [16], [17], [19]	7
L.G2	Attitude towards customer satisfaction and needs	[12], [13], [14], [16], [18]	5
L.G3	Belief that a working product is the basic measure of progress	[12], [13], [14]	3
L.G4	Continuous cooperation with the customer	[13], [14], [4]	3
L.G5	Accurate knowledge of who the customer is and what are their needs	[14]	1
L.G6	Cooperation with the customer based on partnership	[16]	1
L.T1	Mutual trust	[12], [13], [14], [4], [15], [16], [19], [20]	8
L.T2	Direct communication - face to face conversations	[12], [13], [14], [4], [15], [16], [19]	7
L.T3	Focus on achieving common goal	[12], [13], [14], [15], [18], [19]	6
L.T4	Mutual respect	[14], [4], [15], [17], [19]	5
L.T5	Helping each other	[12], [14], [15]	3
L.T6	Taking into account the opinions of other people	[13], [15]	2
L.T7	Respecting the experience and skills in all team members	[13], [14]	2
L.T8	Listening to the opinions of other people	[14], [15]	2
L.T9	Team responsibility	[14], [16]	2
L.T10	Treating team members as people, not a resource	[14], [20]	2
L.T11	Openness to others	[14], [20]	2
L.T12	Sincerity	[14], [21]	2
L.T13	A relaxed atmosphere	[19], [20]	2
L.T14	Equality in the team	[14]	1
L.T15	Sense of security	[4]	1
L.T16	Focus on people instead of on processes	[16]	1
L.T17	Not blaming each other	[16]	1
L.T18	Not covering up the failures	[18]	1
L.T19	Searching for a solution to the problem instead of finding the guilty	[18]	1
L.I1	Continuous improvement and learning	[12], [13], [14], [4], [15], [16], [17], [18], [20]	9
L.I2	Openness to change	[12], [13], [14], [4], [16], [17], [18], [20]	8
L.I3	Being motivated	[12], [13], [14], [16], [19], [20]	6
L.I4	Treating failure as an opportunity to learn, learning from mistakes	[4], [16], [17], [20], [21]	5
L.I5	Creativity, innovation	[13], [18], [19]	3
L.I6	Ability to accept failure and deal with it	[17], [18], [21]	3
L.I7	Taking risks	[4], [17]	2
L.I8	Willingness to constantly acquire knowledge	[15], [18]	2
L.I9	Positive attitude	[18], [19]	2
L.II0	Assertiveness	[14]	1
L.II1	Focus on the task being performed	[4]	1
L.II2	A sense of pride in the job	[17]	1
L.II3	Not giving up	[18]	1
L.II4	Inquisitiveness	[18]	1
L.II5	Pragmatism	[18]	1
L.O1	Self-organization	[12], [13], [14], [4], [15], [16], [19]	7
L.O2	Ability to collaborate	[12], [13], [14], [4], [16], [17], [20]	7
L.O3	Maintaining a steady pace of work	[12], [13], [14], [4], [16], [20]	6
L.O4	Sharing knowledge and results	[12], [13], [14], [18], [19], [20]	6
L.O5	Simplicity and maximization of unnecessary work, simplifying tasks	[12], [13], [14], [4], [16]	5
L.O6	Transparency in decision-making and actions	[12], [14], [4], [20], [21]	5
L.O7	Ability to make decisions together	[12], [13], [14], [15]	4
L.O8	Interdisciplinarity	[12], [13], [14]	3
L.O9	Attitude towards working in short iterations with small increments	[14], [16]	2
L.O10	Applying retrospectives to identify areas for improvement	[14], [16]	2
L.O11	Understanding the purpose and vision of the task before taking it	[4], [15]	2
L.O12	Focus on cross-functional teams	[15]	1
L.O13	Expressing feedback on the work of other people	[15]	1
L.O14	Estimating the results for a given timeframe	[16]	1
L.O15	Determining possible tasks instead of looking for excuses	[18]	1
L.O16	Asking questions in case of insufficient knowledge	[20]	1
L.O17	Focus on one task instead of many at once	[21]	1
L.O18	Finishing the current task before taking the next one	[21]	1

C. Final list merged from literature and interviews

Finally, we merged the lists of agile mindset elements identified from literature and with the interviews. The resulting list of unique agile mindset elements comprises 70 entries, which exceeds the limitations of this paper. However, all identified agile mindset elements were already shown in Table II and Table III. Table IV shows the number of agile mindset elements in each category identified in the literature and the interviews as well as the number of unique elements in our final list.

TABLE IV.  
NUMBER OF IDENTIFIED AGILE MINDSET ELEMENTS

Category	Literature	Interviews	Unique
Support for business goals	6	4	7
Relationships within the team	19	11	20
Individual features	15	16	24
Organization of work	18	8	19
Total	58	39	70

IV. SELECTION OF AGILE MINDSET ELEMENTS

Based on the criteria presented in section II, we have selected 26 elements of agile mindset out of 70 for further evaluation with the opinion survey. We selected 3 elements out of 7 in the “support for business goals” category, 10 elements out of 20 in the “relationships within the team” category, 6 elements out of 24 in the “individual features” category, and 7 elements out of 19 in the “organization of work” category. We could observe that 13 elements in the “individual features” category as well as 8 elements in the “organization of work” category were mentioned only in one source, be it literature or interview.

The list of elements selected for the survey is shown in Table V.  $n_L$  column presents the number of literature sources, while  $n_E$  column presents the number of experts mentioning each element. The final unique agile mindset elements were given new identifiers prefixed with the category symbol only, as described in section III. The identifiers of the merged elements from the literature (see Table III) and the interviews (see Table II) are given in columns  $ID_L$  and  $ID_E$ , respectively.

TABLE V.  
ELEMENTS OF THE AGILE MINDSET SELECTED FOR THE SURVEY

ID	Element name	$n_L$	$n_E$	$ID_L$	$ID_E$
G1	Continuous delivery of a valuable product in short intervals	7	1	L.G1	E.G3
G2	Cooperation with the customer based on partnership	1	3	L.G6	E.G1
G3	Attitude towards customer satisfaction and needs	5	2	L.G2	E.G2
T1	Mutual trust	8	5	L.T1	E.T1
T2	Direct communication - face to face conversations	7	2	L.T2	E.T9
T3	Focus on achieving common goal	6	3	L.T3	E.T7
T4	Helping each other	3	4	L.T5	E.T3
T5	Sincerity	2	4	L.T12	E.T2

T6	Mutual respect	5	3	L.T4	E.T5
T7	Mutual listening	0	3	-	E.T4
T8	Equality in the team	1	3	L.T14	E.T6
T9	Searching for a solution to the problem instead of finding the guilty	1	2	L.T19	E.T8
T10	Team responsibility	2	2	L.T9	E.T10
I1	Continuous improvement and learning	9	3	L.I1	E.I3
I2	Openness to change	8	5	L.I2	E.I1
I3	Being motivated	6	2	L.I3	E.I4
I4	Positive attitude	2	3	L.I9	E.I2
I5	Openness to criticism and feedback	0	2	-	E.I5
I6	Openness to others	0	2	-	E.I6
O1	Self-organization	7	4	L.O1	E.O1
O2	Maintaining a steady pace of work	6	2	L.O3	E.O4
O3	Ability to collaborate	7	0	L.O2	-
O4	Sharing knowledge and results	6	1	L.O4	E.O6
O5	Asking questions in case of insufficient knowledge	1	3	L.O16	E.O3
O6	Finishing the current task before taking the next one	1	3	L.O18	E.O2
O7	Transparency in decision-making and actions	5	2	L.O6	E.O5

V. EVALUATION OF AGILE MINDSET ELEMENTS

A. Characteristics of respondents

The evaluation survey was carried out in late June and early July 2018. The questionnaire comprised 5 sections: an introductory section and 4 sections with the agile mindset elements to evaluate grouped into their categories. In total, 52 respondents took part in the survey. Table VI shows the distribution of the respondents’ experience with agile. Most of the respondents (52%) had at least 2 years of experience.

TABLE VI.  
DISTRIBUTION OF THE EXPERIENCE OF SURVEY RESPONDENTS

Experience years	n
<1	7
1-2	18
2-3	13
3-5	7
>5	7

Table VII shows the distribution of respondents’ roles in agile teams. Most of them worked as developers (about 60%), while others worked mostly as Scrum Masters.

TABLE VII.  
DISTRIBUTION OF THE ROLES OF SURVEY RESPONDENTS

Role	n
Developer	31
Scrum Master	13
Tester	3
Product Owner	2
Agile Coach	1
Analyst	1
UX Designer	1

### B. Evaluation of agile mindset elements and categories

Table VIII presents the evaluation of the importance of selected agile mindset elements to the effectiveness of an agile team according to the respondents' opinion. E shows the mean evaluation of an agile mindset element in the Likert-type scale of 0 to 5 with standard deviation; n gives the sample size. The sample size slightly differs for some elements due to the option to skip an element in the survey. The elements are ordered by their decreasing evaluation.

TABLE VIII.  
EVALUATION OF THE IMPORTANCE OF AGILE MINDSET ELEMENTS TO THE TEAM EFFECTIVENESS

No.	ID	Element name	E	n
1	T9	Searching for a solution to the problem instead of finding the guilty	4.44 (0.79)	52
2	I3	Being motivated	4.44 (0.69)	52
3	T4	Helping each other	4.40 (0.63)	52
4	T7	Mutual listening	4.37 (0.71)	51
5	T3	Focus on achieving common goal	4.29 (0.77)	52
6	I5	Openness to criticism and feedback	4.23 (0.82)	52
7	O4	Sharing knowledge and results	4.21 (0.86)	52
8	T6	Mutual respect	4.11 (0.91)	52
9	T1	Mutual trust	4.10 (0.96)	51
10	T5	Sincerity	4.09 (0.97)	52
11	I1	Continuous improvement and learning	4.08 (1.00)	52
12	O7	Transparency in decision-making and actions	4.08 (1.03)	52
13	O1	Self-organization	4.04 (0.88)	52
14	I2	Openness to change	4.00 (1.02)	52
15	G1	Continuous delivery of a valuable product in short intervals	3.96 (1.04)	52
16	G3	Attitude towards customer satisfaction and needs	3.92 (0.83)	52
17	G2	Cooperation with the customer based on partnership	3.88 (0.97)	52
18	I6	Openness to others	3.88 (0.97)	52
19	O6	Finishing the current task before taking the next one	3.86 (1.06)	52
20	I4	Positive attitude	3.84 (0.77)	52
21	O3	Ability to collaborate	3.81 (0.90)	52
22	O5	Asking questions in case of insufficient knowledge	3.74 (1.06)	51
23	T2	Direct communication - face to face conversations	3.69 (1.26)	52
24	T8	Equality in the team	3.42 (1.28)	52
25	T10	Team responsibility	3.23 (1.31)	52
26	O2	Maintaining a steady pace of work	3.04 (1.34)	52

It can be seen that the top evaluated elements reached the evaluation of about 4.5 out of 5. 14 out of 26 elements reached the evaluation of 4.0 and above. They can be considered the recommended agile mindset elements in our survey. The lowest evaluated elements obtained the score of less than 3.5. However it should be noted that the standard deviation of the evaluations of the last 4 elements is the highest in all our study (about 1.3). Other elements were evaluated with the standard deviation of 0.69 to 1.06.

We have also calculated the mean evaluation of all agile mindset elements in particular categories which is presented in Table IX. It can be observed that "individual features" are evaluated as the most important category. Next is "relationships within the team", followed by "support for business goal". "Organization of work" scored the lowest mean evaluation of all categories.

TABLE IX.  
MEAN EVALUATION OF THE AGILE MINDSET CATEGORIES

Category	E	n
Support for business goals	3.92 (0.95)	156
Relationships within the team	4.02 (1.07)	518
Individual features	4.08 (0.91)	312
Organization of work	3.83 (1.09)	363

### C. Analysis of confounding variables

We have analyzed the respondents' experience and role as confounding variables in the evaluations of agile mindset elements. The results are presented in Table X and Table XI.

TABLE X.  
EVALUATION OF AGILE MINDSET ELEMENTS BY EXPERIENCE

ID	Element name	E <sub>exp&lt;2</sub>	E <sub>exp&gt;=2</sub>	p
G1	Continuous delivery of a valuable product in short intervals	4.08	3.82	0.362
G2	Cooperation with the customer based on partnership	3.88	3.93	0.870
G3	Attitude towards customer satisfaction and needs	3.88	3.85	0.914
T1	Mutual trust	4.20	4.00	0.465
T2	Direct communication - face to face conversations	3.76	3.67	0.795
T3	Focus on achieving common goal	4.40	4.19	0.323
T4	Helping each other	4.44	4.37	0.697
T5	Sincerity	4.28	3.93	0.194
T6	Mutual respect	4.12	4.11	0.973
T7	Mutual listening	4.32	4.42	0.614
T8	Equality in the team	3.48	3.37	0.763
T9	Searching for a solution to the problem instead of finding the guilty	4.60	4.30	0.175
T10	Team responsibility	2.96	3.48	0.158
I1	Continuous improvement and learning	4.12	4.04	0.770
I2	Openness to change	3.96	4.04	0.790
I3	Being motivated	4.56	4.33	0.246
I4	Positive attitude	4.04	4.67	0.083
I5	Openness to criticism and feedback	4.36	4.11	0.285
I6	Openness to others	4.24	3.56	0.018
O1	Self-organization	4.28	3.82	0.057
O2	Maintaining a steady pace of work	3.00	3.07	0.393
O3	Ability to collaborate	4.00	3.63	0.143
O4	Sharing knowledge and results	4.32	4.11	0.393
O5	Asking questions in case of insufficient knowledge	3.96	3.56	0.184
O6	Finishing the current task before taking the next one	3.92	3.93	0.981
O7	Transparency in decision-making and actions	3.96	4.19	0.443

We have used the t-Student test for independent pairs to analyze the differences in mean evaluations depending on experience and role. Treating our data as numerical, this test is suitable for such analysis [11]. We assumed equal variances of the grouped samples and the confidence level of 95% ( $\alpha=0.05$ ).

For the experience test we divided our sample into two groups: less than 2 years of experience and 2 or more years of experience (group sizes were 25 and 27, respectively, which satisfies the prerequisites to the selected test). The mean evaluations are given in Table X in the  $E_{exp<2}$  and  $E_{exp>=2}$  columns, respectively, followed by the p-value of the t-Student test.

For the role test we divided our sample into two groups: developers and non-developers (group sizes were 31 and 21, respectively). Other divisions were not possible due to insufficient number of samples for the prerequisites of the selected test. The mean evaluations are given in Table XI in the  $E_{dev}$  and  $E_{ndev}$  columns, respectively, followed by the p-value of the t-Student test.

TABLE XI.  
EVALUATION OF AGILE MINDSET ELEMENTS BY ROLE

ID	Element name	$E_{dev}$	$E_{ndev}$	p
G1	Continuous delivery of a valuable product in short intervals	3.98	3.91	0.832
G2	Cooperation with the customer based on partnership	3.87	3.95	0.775
G3	Attitude towards customer satisfaction and needs	3.87	3.86	0.959
T1	Mutual trust	4.00	4.25	0.371
T2	Direct communication - face to face conversations	3.39	4.14	0.035
T3	Focus on achieving common goal	4.36	4.19	0.459
T4	Helping each other	4.42	4.38	0.833
T5	Sincerity	4.03	4.19	0.571
T6	Mutual respect	4.19	4.00	0.463
T7	Mutual listening	4.47	4.24	0.269
T8	Equality in the team	3.48	3.29	0.595
T9	Searching for a solution to the problem instead of finding the guilty	4.52	4.33	0.426
T10	Team responsibility	3.19	3.29	0.808
I1	Continuous improvement and learning	3.94	4.29	0.222
I2	Openness to change	3.87	4.19	0.276
I3	Being motivated	4.52	4.33	0.359
I4	Positive attitude	3.84	3.86	0.934
I5	Openness to criticism and feedback	4.32	4.10	0.338
I6	Openness to others	3.94	3.81	0.655
O1	Self-organization	4.07	4.00	0.799
O2	Maintaining a steady pace of work	2.90	3.24	0.388
O3	Ability to collaborate	3.71	3.95	0.350
O4	Sharing knowledge and results	4.19	4.24	0.858
O5	Asking questions in case of insufficient knowledge	3.83	3.62	0.489
O6	Finishing the current task before taking the next one	3.65	4.19	0.070
O7	Transparency in decision-making and actions	4.10	4.05	0.870

Both tests showed that the impact of both experience and role on nearly all of the evaluations could not be considered statistically significant with the assumed confidence level of 95% and sample size of 52. However, two agile mindset elements stood out. The evaluation of I6 element “Openness to others” exhibited statistically significant difference in the evaluation depending on respondents’ experience ( $p<\alpha$  in Table X). It was evaluated much higher (4.24 compared to 3.56) by the respondents with less than 2 years of experience. The evaluation of T2 element “Direct communication - face to face conversations” exhibited statistically significant difference in the evaluation depending on respondents’ role ( $p<\alpha$  in Table XI). It was evaluated much lower (3.39 compared to 4.14) by the developers.

D. Comparison to the principles behind Agile Manifesto

We have mapped the elements of agile mindset in our study to the 12 principles behind the Agile Manifesto [2] and analyzed the evaluation and relative position of the agile mindset elements that map directly onto these principles. The results are shown in Table XII. P# column shows the Agile principle number.

TABLE XII.  
MAPPING OF AGILE MINDSET ELEMENTS ON AGILE PRINCIPLES

No.	ID	Element name	E	P#
2	I3	Being motivated	4.44 (0.69)	5
9	T1	Mutual trust	4.10 (0.96)	5
11	I1	Continuous improvement and learning	4.08 (1.00)	12
13	O1	Self-organization	4.04 (0.88)	11
14	I2	Openness to change	4.00 (1.02)	2
15	G1	Continuous delivery of a valuable product in short intervals	3.96 (1.04)	1, 3, 7
16	G3	Attitude towards customer satisfaction and needs	3.92 (0.83)	1
17	G2	Cooperation with the customer based on partnership	3.88 (0.97)	4
23	T2	Direct communication - face to face conversations	3.69 (1.26)	6
26	O2	Maintaining a steady pace of work	3.04 (1.34)	8

It can be seen that only the I3 agile mindset element mapped to the 5<sup>th</sup> Agile principle was evaluated very high (4.44, position 2). Elements mapped to most of the Agile principles were evaluated in the middle range (4.10 to 3.88, positions 9 to 17). However, the elements T2 and O2 mapped to 6<sup>th</sup> and 8<sup>th</sup> principle respectively were evaluated very low (3.69 and 3.04, position 23 and 26 (last)). Remaining 2 Agile principles mapped to the agile mindset elements that were excluded from the survey.

VI. THREATS TO VALIDITY

A. Threats to construct and internal validity

We have identified and reduced the following threats to the construct and internal validity of this research related to

the interviews and the survey: (a) interview moderator's bias and influence on experts, (b) misinterpretation of the interview outputs, (c) learning and tiring of the survey respondents, (d) forced answers to the survey.

We have controlled the interview moderator's bias and their influence on experts with the structure of the interview. Each interview followed the same protocol (Section II). To minimize misinterpretations, the interviews were recorded, transcribed and thoroughly analyzed while relistening to the recordings, if necessary. The results of each interview have been coded separately and only then merged together.

The survey questions were not randomized to minimize the impact of learning and tiring of the respondents due to the limitation of the Google Forms tool. However, the survey was conveniently divided into 5 sections and contained only 26 evaluation questions. The survey also allowed the respondents to skip the evaluation of a particular agile mindset element when unsure.

### B. Threats to external validity

We have identified the following threats to the external validity of the interviews and the survey: (a) low number of interviewed experts and survey respondents, (b) insufficient experience of interview experts and survey respondents, (c) interview experts and survey respondents as a convenience sample, (d) interview experts and survey respondents sample limited to Polish IT industry.

We have interviewed 5 experts from the industry. The interviewed experts had 2 to 5 years of experience in Agile. We aimed at covering various roles in an agile team and experiences with various agile methods. We engaged 2 Scrum Masters with broad experience (see Table I) Altogether, the input from experts supplemented the list of 58 agile mindset elements from the literature by 22 new elements (38%), which can be considered a substantial contribution (see Table IV).

We have collected data from 52 respondents in the survey, which definitely exceeded the typical threshold sample size of 30 for the choice of the statistical tests [11]. 52% of the respondents had at least 2 years of experience. 13.5% of the respondents had more than 5 years of experience (see Table VI). The respondents represented various roles in the agile team, which covered diverse points of view (see Table VII). Moreover, we have analyzed the impact of the respondents' experience and role as the confounding variables on the validity of our results, which showed marginal impact (Section V.C).

Our survey sample is not statistically random – it is a convenience sample, although we invited the respondents through various channels like personal and business contacts, interest groups, social media, and recommendations. This method provided for a fairly diverse group of experts and respondents with different experience. The experts and respondents used many agile methods such as: Scrum, Kanban, Scrumban, Extreme Programming, SAFe.

The survey was in Polish and possibly attracted most of the respondents among the peers of one of the authors (P. Gaida) working in the Tricity region of Poland, so the results it may exhibit some cultural or regional bias, which needs to be studied further. Comparison of the perception of the concept of agile mindset in Poland and other countries may bring valuable insights.

We have asked our respondents only for their (self-declared) experience in agile and their role in an agile team. We have not collected other data such as company size, age, industry sector or type of projects they worked on. Thus, our study provides only preliminary insight into the conceptual structure of the agile mindset.

## VII. DISCUSSION

The top 5 evaluated agile mindset elements are: "Searching for a solution to the problem instead of finding the guilty", "Being motivated", "Helping each other", "Mutual listening", and "Focus on achieving common goal". They belong only to two categories: "relationships within the team" and "individual features". This suggests that effective agile teamwork requires a specific attitude towards the team and other people as well as proactive and open mind of the individuals. This corresponds with the "growth mindset" concept from Dweck [3].

The 5 least important mindset elements in our survey are: "Asking questions in case of insufficient knowledge", "Direct communication - face to face conversations", "Equality in the team", "Team responsibility", "Maintaining a steady pace of work". They are related to organizational issues as well as shared responsibility and equality. This suggests that agile mindset is not about particular detailed practices or rituals. This is consistent with earlier findings [21, 22, 23, 24].

We have found that less experienced respondents evaluated the "Openness to others" mindset element much higher than those with more than 2 years of experience. Our working hypothesis is that it is related to learning and gathering experience at the start of the professional career. However, "openness" in general is crucial to being agile [2].

The developers considered "face to face communication" less important than the non-developers. Our working hypothesis is that they may see the meetings as (partial) waste of time that diverts them from coding. This may also indicate some overuse or misuse of meetings in the agile teams of our respondents.

We were also able to map the principles behind the Agile Manifesto [2] onto 10 evaluated elements of the agile mindset. These elements occupied positions 2, 11, 13, 14, 15, 16, 17, 23, and 26 (last) in our ranking ordered by the descending evaluation. This is an interesting discrepancy between what our respondents think is important to "being agile" and what the creators of the Agile Manifesto pointed out as the principles of Agile. We can hypothesize that this indicates insufficient understanding of Agile by our

respondents, partial or flawed implementation of Agile in the respondents' teams or companies, or even a shift in practical agility from the 18 years old principles of Agile. This may also be specific to Polish IT industry and have some cultural background. Definitely, it calls for more research.

Our study is based on limited data on the respondents themselves. We have clustered the data by two levels of experience (below and above 2 years) and two types of roles (developers and non-developers). The understanding of the agile mindset may also vary by the industry sector, company size, company culture and maturity, type of projects, national and regional culture and possibly more. Our initial set of agile mindset elements may be used in such further studies.

### VIII. CONCLUSIONS

We have identified 70 elements of the agile mindset from the literature and the industry experts, which answer our research question RQ1. We grouped the elements into 4 categories. Then, we have obtained an opinion-based evaluation of the importance of each agile mindset element to the effectiveness of an agile team, which answers our research question RQ2. Finally, we have analyzed and compared the evaluations to point out the most and least important elements based on the opinions of our respondents, which provides a preliminary answer to our research question RQ3. Further and more detailed study of the impact of agile mindset on the team effectiveness requires careful observation of a number of different types of projects and can be done in future research.

The detailed contribution of this paper is the identification of the elements of agile mindset as well as a preliminary evaluation of their importance to the effectiveness of an agile team based on an industrial opinion survey. This contributes to filling the gap in the literature related to the definition and scope of the agile mindset and the relative importance of its elements in the industry and education [22, 23, 24, 25].

The proposed list of agile mindset elements and their evaluations may be used as a guidance for developers, Scrum Masters and Agile coaches, where the possible applications include: (1) support for the Scrum Masters or coaches in improving the understanding of Agile by the Development Teams; (2) recommendations of improving the agile process and solving problems identified during retrospectives; (3) education and training both in the industry and academia; (4) self-development of the developers, in particular those seeking to switch to Scrum Masters or coaches. Full results of this research are available in [26]. The raw results of our survey are available in [27].

### ACKNOWLEDGMENT

The authors thank all the experts and respondents who took part in the interviews and the survey.

### REFERENCES

- [1] *Manifesto for Agile Software Development*, agilemanifesto.org, 2001
- [2] *Principles behind the Agile Manifesto*, <http://agilemanifesto.org/principles.html>, 2001
- [3] S. C. Dweck, *Mindset: The New Psychology of Success*, Random House, 2006
- [4] G. Broza, *The Agile Mindset – Making Agile Processes work*, 3P Vantage Media, 2015
- [5] A. Przybyłek, W. Kowalski, *Utilizing online collaborative games to facilitate Agile Software Development*, Proceedings of the 2018 Federated Conference on Computer Science and Information Systems, M. Ganzha, L. Maciaszek, M. Paprzycki (eds). ACSIS, Vol. 15, pp. 811–815, 2018, DOI: 10.15439/2018F347
- [6] R. Colomo-Palacios, C. Casado-Lumbreras, P. Soto-Acosta, F. J. García-Peñalvo, E. Tovar-Caro, *Competence gaps in software personnel: A multi-organizational study*, Computers in Human Behavior 29 (2), pp. 456–461, 2013, DOI: 10.1016/j.chb.2012.04.021
- [7] *The 13th annual State of Agile Report*, CollabNet VersionOne, 2019
- [8] K. Schwaber, J. Sutherland, *The Scrum Guide. Rules of the Game*, Scrum.org, 2017
- [9] M. Hammarberg, J. Sunden, *Kanban in Action*, Manning Publications, 2014
- [10] R. Knaster, D. Leffingwell, *SAFe 4.5 Distilled: Applying the Scaled Agile Framework for Lean Enterprises*, 2nd Edition, Addison-Wesley Professional, 2018
- [11] W. Navidi, *Statistics for Engineers and Scientists*, 4th Edition, McGraw-Hill Education, 2014
- [12] J. A. Crowder, S. Friess, *Agile Project Management: Managing for Success*, Springer, 2015
- [13] M. E. Moreira, *The Agile Enterprise: Building and Running Agile Organizations*, Apress, 2017
- [14] M. E. Moreira, *Being Agile, Your Roadmap to Successful Adaptation of Agile*, Apress, 2013
- [15] C. J. Stettina, W. Heijstek, *Five Agile Factors: Helping Self-management to Self-reflect*, A Study of Software Development Team Dynamics in SPI, R.V. O'Connor, J. Pries-Heje, and R. Messnarz (Eds.), EuroSPI 2011, CCIS 172, Springer, 2011, pp. 84–96, DOI: 10.1007/978-3-642-22206-1\_8
- [16] A. Stellman, J. Greene, *Learning Agile: Understanding Scrum, XP, Lean and Kanban*, O'Reilly Media, 2015
- [17] S. McIntosh, *What Exactly is the Agile Mindset?*, InfoQ, 2016, <https://www.infoq.com/articles/what-agile-mindset>
- [18] L. Howard, *What does it mean to have an agile mindset?*, Agile Connection, 2015, <https://www.agileconnection.com/article/what-does-it-mean-have-agile-mindset>
- [19] S. Parvatam, *It's All About the Mindset*, Scrum Alliance, 2015, <https://www.scrumalliance.org/community/articles/2015/december/its-all-about-the-mindset>
- [20] V. Godugu, *Fixed Mindset versus Agile Mindset*, Scrum Alliance, 2015, <https://www.scrumalliance.org/community/articles/2015/june/fixed-mindset-versus-agile-mindset>
- [21] L. Rich, *Agile Is Not a Process, It's a Mindset*, Agile Connection, 2018, <https://www.agileconnection.com/article/agile-not-process-it-s-mindset>
- [22] F. Zieris, S. Salinger, *Doing scrum rather than being Agile: A case study on actual nearshoring practices*, IEEE 8th International Conference on Global Software Engineering, pp. 144–153, 2013, DOI: 10.1109/ICGSE.2013.26
- [23] H. van Manen, H. van Vliet, *Organization-Wide Agile Expansion Requires an Organization-Wide Agile Mindset*, A. Jedlitschka et al. (Eds.): PROFES 2014, LNCS 8892, Springer, pp. 48–62, 2014
- [24] A. Martin, C. Anslow, D. Johnson, *Teaching Agile Methods to Software Engineering Professionals: 10Years, 1000 Release Plans*, H. Baumeister et al. (Eds.): XP 2017, LNBP 283, Springer, pp. 151–166, 2017, DOI: 10.1007/978-3-319-57633-6\_10
- [25] G. C. Gannod, W. F. Eberle, D. A. Talbert, R. A. Cooke, K. Hagler, K. Opp, J. Baniya, *Establishing an agile mindset and culture for workforce preparedness: A baseline study*, Proceedings – Frontiers in Education Conference, 2019, DOI:10.1109/FIE.2018.8658712
- [26] P. Gaida, *Analysis of the agile mindset for software projects*, MSc Thesis, supervisor J. Miler, Gdansk University of Technology, Poland, 2018 (in Polish)
- [27] J. Miler, P. Gaida, Survey on agile mindset and agile team effectiveness, data.mendeley.com, DOI: 10.17632/phvx6nts6b.1



# Scaling agile on large enterprise level – systematic bundling and application of state of the art approaches for lasting agile transitions

Alexander Poth  
Volkswagen AG  
D-38440 Wolfsburg, Germany  
alexander.poth@volkswagen.de

Mario Kottke  
Volkswagen AG  
D-38440 Wolfsburg, Germany  
mario.kottke@volkswagen.de

Andreas Riel  
Grenoble Alps University  
G-SCOP Laboratory  
F-38031 Grenoble, France  
andreas.riel@grenoble-inp.fr

**Abstract**—Organizations are looking for ways of establishing agile and lean process for delivery. Many approaches exist in the form of frameworks, methods and tools to setup an individual composition for a best fit. The challenge is that large organizations are heterogeneous and diverse, and hence there is no “one size fits all” approach. To facilitate a systematic implementation of agile and lean, this article proposes a transition kit based on abstraction. This kit scouts and bundles state of the art methods and tools from the agile and lean community to align them with governance and compliance aspects of the specific enterprise. Coaching of the application of the transition kit ensures an adequate instantiation. The instantiation handles business domain specific aspects and standards. A coaching governance ensures continuous improvement. An example of the systematic application of the transition approach as well as its scaling is demonstrated through its application in the Volkswagen Group IT.

## I. INTRODUCTION

THE DIVERSITY of an enterprise’s business areas demands individualized implementations of lean and agile. Often the main goal of the agile transition is to gain delivery speed. According to Albert Einstein: “Make everything as simple as possible, but not simpler”, we have to find a way to achieve effectively the simple yet complete organizational setting. Furthermore, Conway’s law [44] leads us to develop something customizable to build a lean and agile organization for a best fit to the specific products and services, which the organizational unit creates and delivers. These two aspects have to be handled to realize a lasting and sustainable transformation.

Large established enterprises are built around different business areas with independent business units or divisions in a matrix structure [1]. Most of these business units have the size of a medium-sized enterprise. Furthermore, large enterprises are mostly based on large delivery pipelines oriented on the efficiency paradigm of the Taylorism [45]. Any transition aid for application within such context has to be able to handle this setting. More specifically during our first operational coaching of projects within the Volkswagen Group IT in past transformation initiatives we identified the following aspects an agile transition aid has to address:

- 1) Identify the target organization for the transition, including its boundaries.
- 2) Identify the organization’s value stream, including interfaces at the boundary to “external” partners.
- 3) Define and clarify the transition’s objectives.
- 4) Evaluate different approaches to lean and agile for their suitability in the particular organizational context.
- 5) Implement the selected approaches:
  - Train people in the approach.
  - Re-organize the workflows according to the approach.
  - Align the new setting with the enterprise’s governance and compliance structures.
- 6) Install cyclic checks for transparency and improvement:
  - on a local view of the transition for “self-optimization”;
  - on a global enterprise view to develop the “setting”;
  - offer an open networking platform to reflect transitions.
- 7) Support scaling of transitions

This leads to the investigation question: How is it possible to address these demands with an easy to handle approach, which can be applied by a team of coaches in a structured fashion? Our objectives for achieving this are the following:

- (O1) A transition kit is needed that is able to handle lean and agile approaches.
- (O2) Based on the organization’s stakeholders’ current mindsets a specific set of methods and tools for the workflows has to be implemented.
- (O3) The organization’s specific product setting has to be taken into account appropriately.

## II. RELATED WORKS

This section investigates related published work with a focus on a holistic approach to addressing those. There is a huge amount of relevant approaches to organizational development [2], alternative setups like holacracy [3] or transitions [34] starting on grounded theories [32] to practice collections of other enterprises [33]. We are interested in identifying well-known approaches, methods and tools that can be used as a kind of reference in various settings to reduce complexity. Our contribution is to bring together the

team setting with its cultural and mental history thanks to an adequate set of approaches, methods and tools to realize a effective and sustainable transition. We structured related work according to this scope, rather than elaborating on all kinds of available methods and tools at the time of writing this article.

#### A. Setting Analysis

The Cynefin [5] and the Stacey-matrix [7] are approaches to classify the product context into a complexity setting and the drivers of the transformation [36]. This are useful approaches to identify the development context of the transitions product environment. The spiral dynamics model [4] and the Group Development Questionnaire (GDQ) [8] classifies the maturity of a group of humans who focusing together on an objective or purpose. As setup point on the teams maturity for transition approaches and methods this is crucial. Value-stream mapping [6] is an approach to optimize processes in a given setting especially for software [35] which come for the production [46] to the software development [47].

#### B. Lean and Agile Approaches

Scrum [13] and XP [15] are team approaches focusing on agile working. Kanban [14] works in a team and in bigger organizations. SAFe [9], LeSS [10], Nexus [11] and Scrum@Scale [12] are approaches to handle the synchronization of more teams in a bigger organization. Furthermore a lot of variants are existing like Disciplined agile delivery (DAD [48] or Agile modeling (AM) [49].

#### C. Methods and Tools

Design Thinking (DT) [16] is a method to develop an initial product in an iterative hypothesis based manner. Minimum Viable Product (MVP) [17] and derivations like Simple Lovable and Complete (SLC) [18] are tools to define an initial product version for delivery. Business Model Canvas (BMC) [19] or Lean Canvas [50] and its variants like for organizations internal communication [20] are used to identify the setting of a business to optimize in a later step the value-stream for product and its revenues. The Product Vision Board (PVB) [21] is used to for focusing a team on a product. INVEST [22] is used to systematically identify requirements for a product. Definition of Done (DoD) [23] or derivations like Levels of Done (LoD) are used to ensure that product versions fit quality definitions. To keep the delivery procedure lean and focused Product Quality Risk (PQR) [24] mitigation can guide to the delivery.

#### D. Organizational culture and team psychology

The culture moves to a more internal lean start up [26] setting also in bigger enterprises. The objective of most digital business models [19] is scaling into the mass-markets [25]. Coaching approaches are reflected to be effective in the setting [27] to address the agile teams.

#### E. Governance, Risk and Compliance

Governance has to establish standards like ISO 9000 for quality management, standards for risk management like the ISO 31000 and additional domain specific standards. Approaches for agile risk handling exists [31]. For service management, the ISO 20000 is an established anchor. Some concepts for agile governance [28] and [29] exist, however their scope is limited to applying agile or lean principles outside a globally acting [30] enterprise context.

### III. TRANSITION PROCESS

Within Volkswagen Group IT, we do not use one given method, model or tool because the organizations' s size demands context adequate approaches. More than 2000 internal employees and a lot of divisions and organizational units indicates the complexity which the transformation has to deal with. Therefore we decided to start with the basis: the team.

A transition kit and process has been developed and maintained by a central team, the Agile Center of Excellence (ACE), which guides and coaches agile transitions. ACE is a department within the Group IT uniting initial agile users from the first agile projects. The transition process consists of three phases: the transition itself, as well as a pre- and post-transition phase to ensure sustainable transitions. ACE supports transitions in the Group IT and other business areas of the Volkswagen AG based on their transition process and kit that has been enhanced over years.

The coaches establish the initial setup and alignment during the coaching phase of the team's external process expectations (figure 1). This is the initial link to process safety and compliance for the teams. The long-term alignment is checked by the project review.

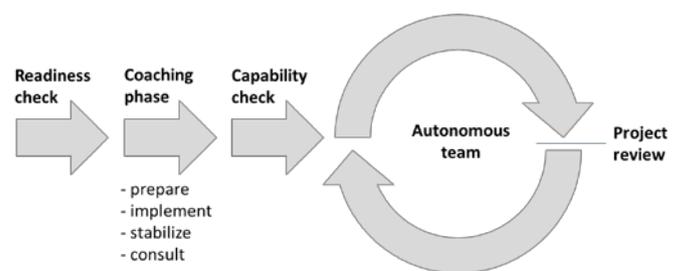


Fig. 1: Coaching to team autonomy with integrated compliance check

In the pre-transition phase, the “readiness check” is conducted to identify the status quo and objectives of the transition. The status quo identifies roles like sponsor of the transition, product/business owner and the team setting. Furthermore, agile artifacts like for the backlog and its items are investigated. Based on the evaluation of the acquired information, a transition can be recommend or not. In case of recommendation, the ACE can support the transition with the transition kit. In case of a non-recommendation to start a transition, the ACE will not support because there is low

chance to finish the transition in time successfully. The biggest challenge during this phase is the interlocutor's honesty. All transition aspects are based on it and conveying information honestly and completely is needed to give the transition a chance to be successful. Therefore we decided that we start the transition with motivated and voluntary units supporting the transition and meeting the prerequisites from the outset.

The main purpose of the transition kit is to enable teams to deliver most product benefit within in a continuously changing environment. The ACE coaches help start agile projects and teach the team how to deal with impediments. Additional ACE tasks are:

- first aid in network,
- promoting agile methods,
- connecting committees,
- supporting knowledge transfer,
- combining agility practices of brands,
- enable leadership to act in an agile way,
- sensitize the unit to get an agile mindset,
- pay attention to process safety.

Every transition phase starts with a contract clarification to get a clear understanding of what will happen. Referring to the Agile Manifesto [39], the contract does not describe the HOW, but rather the WHAT. Depending on the results of the "readiness check" and the needs of a team, product or project, the transition duration will be estimated and a coaching package will be offered (cf. Section V). The contract defines the purpose, deliverables from both sides and the organizational issues like contractor and cost issues. The transition itself has four steps:

1. Preparation (evaluation of team and product setting)
2. Implement the methods and the tooling
3. Stabilization
4. Consulting

The *preparation* includes the execution of a kickoff workshop, consulting (project leads, development team) and agile workflow creation. Also includes support, moderation, preparation of the management and creation of Definition of Ready/Definition of Done and initial product backlog with the team. The initiation of the first meetings like refinement, planning, review and retrospective is a task, too.

To *implement the methods and the tooling* the guide is always available for the team. The coaches train the team and the roles inside e.g. Scrum Master, Product Owner etc. to do the job to be done. The guide also moderates the necessary meetings like review, daily, retrospective, planning or refinement. Furthermore the guide assists the change management for motivation, conflict solving and workflow changes. The coaches are instantiating the initial setup and alignment of team external process expectations. This is the initial link to process safety and compliance for the teams. The long-term alignment is checked by the project review of the post-transition phase.

The *stabilization* step during the coaching (figure 1) is not so intensive for the coaches because the team should do their

first steps alone. The coaches are always available for support and assistance, and in special cases will also assume the role of moderators. In this step, their job is to motivate, inspect, adapt and strengthen the change to be sustained. Solving conflicts is also part of it.

The *consulting* step is demand driven and mostly the end of the transition phase (figure 1). If the customer needs help, the coaches will help and give answers for questions to events, roles and workflow. The guides help the change management manage conflicts and adapt innovations.

The post-transition phase starts with a hold back (capability check in figure 1) of the transition team during the stabilization step and ends with a report. The report reflects the coaching contract objectives and also the agile issues and elements. Furthermore, the team or organization is registered as "agile". This flag will be used for the future agile governance checks (cf. Section VI) to ensure sustainability of the transition and incremental development of the people to stay up to date about the state of the art about agile.

#### IV. TRANSITION KIT

For the demand of the Volkswagen Group IT to transform classic project management to business agility we developed the transition kit. It contains the methods and tools which are released during the transition process. Within the transition process, we try to find the best choice of approaches, methods and tools to create value faster. The transition kit addresses the implement step of figure 1. The transition kit focusses on the key parts of figure 2. These key parts are the product or service which is the delivery to the customer, the team realizing and supporting the products, as well as the governance ensuring organizational standards. Governance can also be triggered by external demands for example from legislative changes. The transition kit has to support the setup of the demanded skills and capabilities of the team from the outcome view (product/service). Furthermore the governance has to handle the product or service risks by guiding the teams to be able to balance the business value and risks related to the product or services they handle.

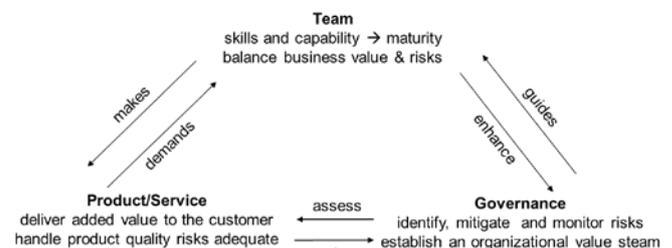


Fig. 2: Transition's key parts and their relationships

All three parts interact and need a holistic handling by the transition kit to realize a comprehensive product or service from the customer view who is using the product/service. The tool selection of the transition kit (table 1) is initially based on a first fit for purpose. This first fit was realized by a literature review [40] to identify artifacts for the initial transition kit. The transition kit contains approaches, methods and tools which helps the coach and team to go in an effective way into the right direction during the transition. Over the life cycle the transition kit will be enhanced by adding and changing artifacts to better fit the current organizational culture, for an easier integration into the coaching or simpler use in a self-service approach for teams without coaches. The enhancement is triggered by feedbacks. While everybody can suggest new artifacts for the transition kit, the ACE will evaluate and integrate relevant suggestions during their cyclic inspections. The objective is not to have a maximum of possible elements in the transition kit, but rather to have a lean transition kit that can be trained easily and is effective in most organizational settings. To make it easy to find the right artifacts the transition kit is aligned with the product complexity, team maturity and the agile approaches.

To identify the projects the ACE supports with coaches we use the Stacey matrix. It is an easy to use way to identify if agile is helpful or not.

The assignment of tools to phases is based on experience during the supported transitions. The determination of the appropriate transition kit artifacts is done according to the

following procedure: To start in a value-driven way, the initial focus of the transition is the product or service. The product is located on the Stacey-matrix. Over the product life-cycle, the complexity location is more or less stable in emerging markets – with a trend to reduction of complexity in mature markets or at the end of a product life-cycle. The current state is identified and the future result or objective will be considered to advance in the right direction. In a second step, the relevant governance guidelines are identified. Based on the product and governance demands, the current team skills and capabilities are focused on. The product team setting is located in the spiral dynamics model (table 2) color levels. This location is important because often organizations coined by Taylorism established over years, act on the “red level”. These teams have to make their mindset leaner to achieve the “blue level”. Agile teams typically act on levels of blue and higher. Each team has to grow level by level in their maturity. This leads to the adaptation of the used artifacts over the maturity journey of a team. Based on the team’s maturity and their product environment complexity, the appropriate agile approach will be selected mostly based on the suggestions of table 2, however the guide and the team can make adjustments if they think another artifact would fit better. The artifacts help the team to progress in the transition, but most of the transition effort is to enable and coach the team to deliver a product. Some examples about the experience-based labeling of the table: Why is Kanban applicable in beige teams? Kanban does not define a set on rituals like retrospectives from Scrum which demands a minimum level of trust in the team

TABLE I.  
TRANSITION KIT ARTIFACTS AND THEIR MAPPING TO TRANSITION SPECIFIC KEY-ASPECTS

Method/tool	Spiral dynamics team maturity	Stacey	Phase (average)	Application
Retrospective	Purple or higher	All	pre, mid, post	High (over 75%)
Design Thinking	Blue or higher	All	Pre	Low (under 25%)
Minimum Viable Product (MVP)	Orange or higher	Complex & complicated	Pre	Mid (25% to 75%)
Simple Lovable and Complete (SLC)	Blue or higher	Complex & complicated	Pre, mid	Low
Business Model Canvas (BMC)	Purple or higher	Complex & complicated	Pre	Low
Product Vision Board (PVB)	Purple or higher	Complex & complicated	Pre	Low
INVEST	Purple or higher	Complex & complicated	Mid	Mid
Definition of Ready (DoR)	Blue or higher	All	Pre, mid	Mid
Definition of Done (DoD)	Blue or higher	All	Pre, mid	Mid
Levels of Done (LoD)	Blue or higher	Complex & complicated	Mid	High
Product Quality Risk (PQR)	Red or higher	Complex & complicated	Mid	Low
Scrum	Purple or higher	Complex & complicated	Pre, mid	Mid
Extreme Programming	Green or higher	Complex & complicated	Pre, mid	High
KANBAN	Beige or higher	Complex & complicated	Pre	Low
SAFe	Red or higher	Complex & complicated	Pre, mid	Mid
LeSS	Blue or higher	Complex & complicated	Pre, mid	Low
Nexus	Orange or higher	Complex & complicated	Pre, mid	Low
Scrum@Scale	Orange or higher	Complex & complicated	Pre, mid	Low

TABLE II.  
MATURITY LEVELS OF THE SPIRAL DYNAMICS MODEL [4]

Name	Structure	Motives	Characteristics
Beige	Loose bands	Survival	Archaic, instinctive, basic, automatic
Purple	Tribes	Magic, Safety	Animistic, Tribalistic, Magical, Mystical
Red	Empires	Power, Dominance	Egocentric, Explorative, Impulsive, Rebellious
Blue	Pyramidal	Order, right & wrong	Absolutistic, Obedient, Purposeful, Authoritarian
Orange	Delegative	Autonomy, achievement	Materialistic, Strategic, Ambitious, Individualistic
Green	Egalitarian	Approval, Equality, Community	Relativistic, Personalistic, Sensitive, Pluralistic
Yellow	Interactive	Adaptability, Integration	Systemic, Conceptual, Ecological, Flexible
Tortoise	Global	Compassion, Harmony	Holistic, Global

to discuss issue frankly. Kanban itself is a more “mechanical” approach. Both approaches can be used to develop the teams to higher levels. With higher levels the teams are acting different within the same approach by discovering more opportunities with the higher team trust and openness. Why do we have small “item” like MVP and “big items” like Safe in the table? Depending on the context it is useful to start with small items to support individual transitions of teams. In case of a more top-down demand a big item reduces discussions about how to start because it is like a pre-defined “package” ready for rollout. This is also the reason why the transition kit does not add every approach, method or tool – it selects some (first fit algorithm based) which work in the industrial context and tries to reduce redundancy were it is useful and possible by offering enough variance for the individual coaching of teams. The objective for the transition kit is to offer a practicable way for the transition of a team, without proposing any way possible.

The transition kit does not focus on finance procedures of the enterprise however some programs are using for example MVP based finance planning to manage their annual budgets in an agile fashion. However the approaches, methods and tools can be applied to special functions. For example, the Group IT security organization was an early adapter.

The transition kit is designed to develop culture, team maturity and products/services together. Of course it is possible to enforce some methods or tools on lower leveled teams, but the real opportunities are only realized within the right culture and team context. The application column in

table 2 shows a current distribution of the application the line in teams.

## V. COACHING

ACE offers different volumes of coaching packages [37]. The package size is defined by the amount of time a team gets support from the transition team. The intensity depends on the time the guides (coaches) support the team. The coach sets up the team to address the demands and objectives of the transition by using the transition kit as guidance framework for the transition. The main focus of coaching is on the events, mindset, team performance, roles and their tasks, the used methods and how to inspect and adapt. Therefore the guide will use workshops with the whole team, as well as direct coaching.

Every coaching starts with a collection of information. This is necessary to find out what the transition (e.g. the project or team) really needs. To implement agility, the coach starts creating awareness of agile principles and values. With growing understanding, the flow will be created to support agile behavior. This means that the team can welcome and handle requirement changes having influence on the actors. The coach helps to give the team the power and knowledge they need. This is an ongoing process during all transition phases and may not be finished when the coach leaves the team.

When the transition goal is clear, the coach has to decide on which level to be most effective. If the transition has most effect on teams, the coach will focus on team members. The objective of the coach is to start small and establish the simplest possible set of artifacts from the transition kit to realize the objectives of the transition. For instance, if the coach decides implementing Scrum, he will support the Scrum team including the Scrum Master, the Product Owner and the development team. If the transition requires an organizational change, the coach will spend more time on management level where the responsibility for the portfolio is located. The tools and methods are all based on values and principles. The coach’s main task is to make clear what the effects of their actual application are. Furthermore, the coach facilitates the teams with methods and tools for generic product and service development. An example is requirements elicitation and engineering with the product vision board to align the requirements at least with epics and stories oriented with INVEST and PQR (cf. table 1).

## VI. GOVERNANCE

Each enterprise needs a governance structure ensuring that fundamental things are done in a deterministic way, and at minimum according to the state of the art. The state of the art is defined by organizational settings or derived from the market standard and regulations. Consequently, also all agile and lean teams have to establish and ensure the state of the art for their products and services. Depending on the product

specific aspects, on top of the state of the art additional factors have to be ensured, e.g. market advantages. During the coaching phase aligned with the transition kit this is delivered by a team external coach. The coach has to make the teams sensitive for this governance topic and their team responsibility to stay aligned in the future. After the coaching phase the teams are independent and have to care about the “update” to the developing state of the art on their own. To make it easier for the teams, the governance offers update information about state of the art changes, which can be adopted by the teams. However, the governance has to ensure the alignment with the rail guards and update them to fit the state of the art. Rail guards are typically artifacts ensuring that some basics are done by the teams like for example an approval evidence for a deployment. Furthermore, the governance has to verify the effectiveness of its settings. These effectiveness checks are realized with controls. Different (domain) standards for System and Organization Controls (SOC) like [42] exist, but all have in common that the effectiveness of the established procedures has to be adequately checked, and if needed an alignment action has to be triggered. To ensure alignment with the settings and the agile and lean mindset a project review is established [38]. The project review (see figure 1) checks different aspects of an agile team or organization. Depending on the project or product classification (based on risk etc.) it will be checked in a deterministic way or randomized picked for a review. This ensures a basic transparency of alignment with the state of the art of the current portfolio.

The reviews are conducted by some coaches who have been trained in the evaluation aspects and their rating criteria. This common understanding about the aspects and rating ensures comparable results to derive organizational issues. Furthermore, an objective is not to change existing review aspects to keep the historical results in the data-analysis pool.

The defined rail guards for the expected artifacts and outcomes for fulfilling external requirements like aspects of the GDPR [43] or quality standards like ISO 9000 are checked in the project review. The results are used on both levels, for the reviewed team as well as the overall organization. Most of the findings have to be addressed by the product teams, however some findings are seen in many teams. This is made by cyclic analysis of the project review results to identify “derivation pattern” which have to be addressed on the organizational level. A derivation pattern is identified if in a significant amount of the cyclic checks similar derivations are observed. This is the trigger to handle it not only on the specific product or service instance and start caring about it on a generic or organizational level. For each identified derivation pattern the governance checks why it does not fit to the product teams and their deliveries. This can lead to actions on the organizational level having a high bandwidth. Finally, there is the educational aspect that leads to inadequate setting – this is addressed by training or

coaching offers to establish the things as intended. This may lead to refactoring the rail guards or artifacts to fit better into the project teams and the organizational culture. Figure 2 shows the relation between the product, the team and the governance. The relation “enhance” in figure 2 leads to the learning that as much as possible should be structured as self-service for the teams to reach higher autonomy and better scaling. This initial higher effort to develop the governance outcomes as self-service capability empowers the teams to live their self-organization and responsibility. To give feedback to the teams in a gamification context, the top ranked project review results are posted on an intranet page as a “champions league table” involving the entire organization.

The development and update of the transition kit is an additional important task to assure alignment with current regulations and the developing state of the art over the time. The transition kit has to support the governance artifacts like the rail guards during the team settlement. To do this, external and internal triggers are established. For example, the PQR method from the transition kit directly helps to make transparent why things are done in this way for some governance measures. An objective of the improvement of the transition kit from the governance perspective is to integrate as many measures as possible into the product or service artifacts or their direct production procedures. This integration makes it leaner and easier for the product teams to align their work with the expected outcomes and measures.

The Volkswagen Agile Community (AC) is the chance for everybody to get updates and the information about current development of agile and lean. It is an open community for networking and share knowledge about agile and lean. This includes also topics about the transition kit and agile governance.

DACH30 [41] is a trans-enterprise network to share experience about agile and lean. Trainings and skills are developed together. This ensures that the transition kit is reflected by external experts and is updated to the current insights of other enterprises.

The objective of the governance is to give the teams as much freedom for agility as possible while still demanding sufficient discipline from the teams to fit the compliance framework.

## VII. EXPERIENCE REPORT

At Volkswagen AG Group IT, the transition kit development started in 2016 to support the coaches’ daily work and has been enhanced continuously by the ACE and the coach guild to address the challenges of migrating to lean and agile methods in a structured way. Currently more than 100 product/service teams and organizational entities have been coached based on the elements of the transition kit. All those elements have been deployed – some more often than others (see table 1, column application). The teams are from

the Group IT as well as other areas of the Volkswagen AG like plant production planning or vehicle development organizations, as well as smaller organizations like board member offices. The teams are supported during the transition in different life-cycle phases of their products and service. Some teams started on a green field, some were already established delivery teams. The range of software developed by the coached teams covers a wide range – from standardized ERP systems supporting human resources and production logistics to special software for supporting specific intellectual property of a business area. Also the architecture differs from established 3-tier architectures to cloud native micro-service based systems. The coaching phase differs in time from a few weeks to many months – depending on the size of the team or organization. Additionally, within the Volkswagen AG there exist a number of self-service based transitions which are often unknown to the ACE. By using the self-service, the teams have a low entry barrier because they can do it on their own way and speed, but the risk of applying inadequate elements of the transition kit is higher without an experienced coach.

The following parts of the case study reflect the objectives O1 to O3 and the observations of the application of the transition kit in the coaching phase as well as the results of the project reviews to have a long term perspective on the sustainability of the transition.

The lean and agile approaches are mapped to the transition kit artifacts to support the artifact selection. Depending on the approach, more or less options are offered to be chosen by the coaches and teams (O1). There is a trend in smaller teams without an end to end responsibility to use Kanban. This is motivated by the external process dependencies which limit the team's autonomy and freedom. The teams are often part of process driven value chains which drive the cycle time and delivery-dependencies. Hence, sprint commitments are not easy for the team. On the other side there is a trend to SAFe for transitions of multi-team organizations. Both show that the upper maturity levels are often not achieved.

The maturity derived from the spiral dynamics model of the teams is mapped to the transition kit artifacts to support especially lower leveled teams by choosing adequate approaches. With higher maturity levels the transition kit gets less importance because the teams have the capability of improving on their own and develop their appropriate way with supporting methods and tools to address their specific situation best (O2). Many teams have started their transition from the red or blue level Taylorism driven culture. However, some teams are built from scratch and in a greenfield area. Here, a quick move to “higher” levels is possible, because they do not have to learn to forget established habits and culture. The coaches typically can see some progress of one or two levels during their supporting phase. In the project reviews after a longer time a further progress can be observed. But in case of no strict application

of agile methods and mindset some teams also go down to their “roots” with Taylorism habits. For these teams a “refreshing” coaching phase is suggested, if they still want to become agile.

The specific product setting with the complexity and value stream is supported by the transition kit, too. The artifacts are mostly generic and fit to the typical product settings in the complex setting (O3). In the future it could be possible to simplify the transition kit more by substituting complexity specific artifacts by generic ones.

The fact that the agile teams investigated in the case study are not permanently co-located does not significantly impact the application of the transition kit because most of the teams have some cyclic common physical meetings like refinements or retrospectives and use in-between communication tools to setup virtual team rooms.

The case study identifies that all phases of the transition are applied and supported as intended by the transition kit as described in section IV. The transition kit makes it easier for new coaches to deliver transition support in a project-style to the teams in a standardized way. The integration of the transition kit in the holistic enterprise environment with a centralized product delivery process compliance helps the coaches and teams to be effective also from a compliance perspective. The controls of the effectiveness work because some transitions were not started because the environment did not fit according to the results of the readiness check.

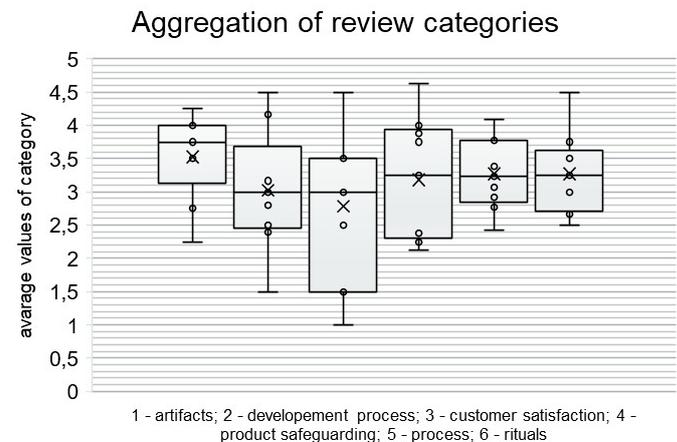


Fig. 3: Anonymized review results of the categories shows spreads and potentials (1 is most left bar – 6 most right bar)

The control project review with its check aspects helps to show the effectiveness of the transition and its sustainability in the teams later on (see Fig. 3). Based on these measurements and metrics for agile projects, agile processes, and agile teams the governance identifies improvement potentials. For example, one related to the agile development process (which is the 2<sup>nd</sup> bar in figure 3) effectiveness controls the re-thinking of the Group IT development process for a better alignment with agile and lean approaches and setting of guide lines which can easier integrated into operational excellence by the teams was indicated.

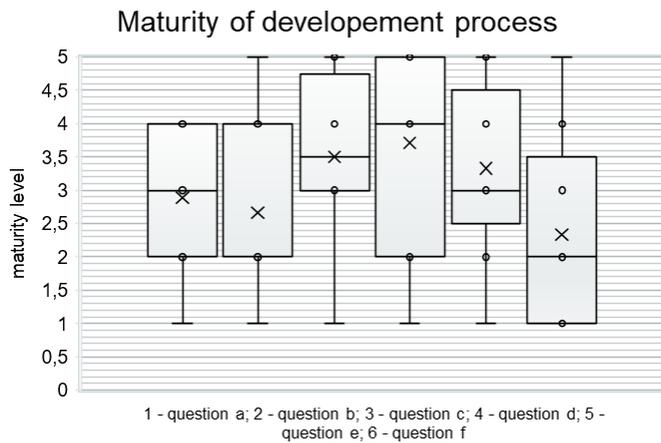


Fig. 4: Maturity of the agile development process (1 is most left bar – 6 most right bar)

Figure 4 shows the results of category agile development process of representative project reviews between 2017 and 2019. The x-axis are checked aspects of the project review which is aligned the teams agile adaption and the governance aspects. A more detailed description of the aspects and their grouping on the x-axis is in [38] described. The y-axis shows the fulfillment of the checked aspect. The bar in the middle shows the 2<sup>nd</sup> and 3<sup>rd</sup> quartile of values. The trend on derivations to the standardized templates of the development process is visible (every question has low values and almost all also high values – especially question f in figure 4). This derivation has led to the creation of a community of practice as a kind of working group whose mission is to enhance the Group IT development process to be better aligned with the state of the art habits of agile and lean working teams. This is one way of feedback to improve the environment to be more agile.

Often the coaches also identify new approaches, methods or tools which are evaluated as a kind of experiment during a selected team coaching. Results and lessons learned from this experiments are reflected in ACE to improve the transition kit. Furthermore, the case study shows that some transitions are not lasting or sustainable. The effectiveness of the transition is checked by the review with a delay to the coaching phase. By comparing the results achieved during the transition with the results of the progress reviews the progress or back-steps of the teams can be made transparent and thereby used for deriving the appropriate improvement actions. The selection of the reviews was made from feedback applications by randomized picking from the successful team transformation list and high-risk labeled projects/products. The highest frequency is one year for conducting reviews in a team. This is to avoid too many reviews in short time periods by random picking without the chance for the teams to improve in between reviews.

## VIII. CONCLUSION

The presented holistic scaling approach demonstrates that a centralized agile governance can help large enterprises

scale agile transitions in the product and service teams. This centralized ACEs coach guild and Agile Community are used to manage the agile knowledge and enhance the transition kit. The setup of a self-service driven team governance is a chance for establishing a lean governance approach. Furthermore the lean and agile mindset in governance offers the teams the chance to participate in the future “look and feel” of the governance, such as the development of higher automation of governance tasks and their evidences. This automation objective is a logical consequence of the automation with the everything as code approach [51] of devops. The governance will check the effectiveness of the participation driven development with the controls like the governance initiated reviews to ensure that the enterprise enhance in a positive way aligned with the strategy. A second observation is that the governance develops fast if they live the lean and agile mindset themselves. Their responsibility is to serve the teams in an effective way to be compliant with external and internal requirements.

The evaluation about the effectiveness of coaching with a transition kit is seen on two points:

- At the end of the coaching phase on which the readiness check situation and the current outcomes of the capability check are compared.
- At the project review with the distance view (at least 1 year) after the transition coaching.

The objective of the ACE is to be effective by the coaching support. This is realized with the transition kit by applying and enhancing the transition kit continuously with the lessons learned from the transitions coaching. The efficiency is seen on the higher team transformation throughput of coaches. The issue is to have a generalized kit which is easy to instantiate in the specific team setting. This trade-off is a current enhancement focus of the transition kit. Furthermore a contribution is that this transition kit explicitly handles the mental team setting by application of the spiral dynamics model to apply adequate approaches and methods during the transformation to support effectivity the progress and sustainability also after the coaching phase.

## IX. NEXT STEPS AND FUTURE WORK

Sustainability is a topic that needs more focus. Often the agile project review makes transparent that after the coached transition phase, the teams lose some of the leaned rituals etc. and fall back to pre-transition habits. We need to define or develop external triggers to reflect the team’s rituals and progress in the developing of the agile and lean mindset without the coaches. This is a topic for an effective governance of the agile and lean processes.

Furthermore, the amount of skilled coaches does not scale with the demand. We need to enhance the transition kit to a complete self-service approach. Then teams with some “basic” skills can work more autonomously, needing less coaching. This is a governance and training issue. The

training aspect is to enable the teams to do mostly everything in a self-service manner by offering a suitable transition kit. But on the other side the governance has to ensure that also self-service transitions have high quality outcomes.

Another open point is that the presented approach is only applied in a European enterprise culture. Its effectiveness in other cultural contexts still has to be investigated.

Next steps are the refactoring of the current process governance rail guards for a higher automation degree. The objective of the potential automation offers mature teams the integration into their automated product delivery pipeline (CI/CD chain). Some teams are currently experimenting and evaluating automated governance controls. The challenge is to find a balance between integrated standard tools and the freedom of the agile teams. Is automation an adequate indicator to determinate the product team maturity, especially in team's customized CI/CD chains? Will an individualized CI/CD chain slow down the integration of currently "independent" agile teams in future release trains of SAFe? Another interesting point is to extend the product based focus of the transition kit with a more lean and agile product finance scope like Beyond Budgeting [52].

#### REFERENCES

- [1] Marvin R. Gottlieb, "The Matrix Organization Reloaded: Adventures in Team and Project Management," Praeger Publishers, 2007, ISBN 0275991334
- [2] Kesler, Gregory, "Leading organization design : how to make organization design decisions to drive the results you want," Kates, Amy. (1st ed.) 2011, pp. 9–10. ISBN 9780470912836
- [3] B. J. Robertson, "Holacracy : the new management system for a rapidly changing world," 2015, Henry Holt & Co, ISBN 9781627794282
- [4] D. Beck, C. Cowan, "Spiral Dynamics: Mastering Values, Leadership, and Change," 1996, Wiley-Blackwell, ISBN 1-55786-940-5.
- [5] C. F. Kurtz, D. J. Snowden, "The new dynamics of strategy: Sense-making in a complex and complicated world," IBM Systems Journal. 42 (3): 462–483, 2003, doi:10.1147/sj.423.0462
- [6] M. Rother, J. Shook, "Learning to See: value-stream mapping to create value and eliminate muda," Brookline, Massachusetts: Lean Enterprise Institute, 1999, ISBN 0-9667843-0-8.
- [7] R.D. Stacey, C. Mowles, "Strategic Management and Organisational Dynamics," Pearson, 2015, ISBN-13: 978-1292078748
- [8] G. Buzaglo, S. A. Wheelan, "The Group Development Questionnaire: A Scientific Method Improving Work Team Effectiveness," Annual Quality Congress, Vol. 51 No. 0 pp. 737-741, 1997
- [9] D. Leffingwell "SAFe® 4.0 Reference Guide: Scaled Agile Framework® For Lean Software and Systems Engineering," Addison-Wesley Professional, 2016, ISBN 978-01234510545
- [10] <https://less.works/>
- [11] <https://www.scrum.org/resources/nexus-guide>
- [12] <https://www.scrumatscale.com/scrums-at-scale-guide/>
- [13] K. Schwaber "Agile Project Management With Scrum," Microsoft press Redmond, 2004, ISBN 978-0735619937
- [14] T. Ohno, « Toyota Production - beyond large-scale production, Productivity Press, 1988, pp. 25–28, ISBN 0-915299-14-3
- [15] K. Beck, "Embracing change with extreme programming," Computer, 32 (10), 1999, pp. 70-77
- [16] J. Liedtka, "Designing for Growth: A Design Thinking Tool Kit For Managers," New York: Columbia University Press, 2011, ISBN 0-231-15838-6
- [17] [www.syncdev.com/minimum-viable-product/](http://www.syncdev.com/minimum-viable-product/)
- [18] <https://blog.asmartbear.com/slc.html>
- [19] [http://www.hec.unil.ch/aosterwa/PhD/Osterwalder\\_PhD\\_BM\\_Ontology.pdf](http://www.hec.unil.ch/aosterwa/PhD/Osterwalder_PhD_BM_Ontology.pdf)
- [20] <https://eee.do/internal-communication-canvas/>
- [21] <https://www.romanpichler.com/tools/vision-board/>
- [22] L. Buglione, A. Abran, "Improving the User Story Agile Technique Using the INVEST Criteria," Joint Conference of the 23rd International Workshop on Software Measurement and the 8th International Conference on Software Process and Product Measurement, Ankara, 2013, pp. 49-53. DOI: 10.1109/IWSM-Mensura.2013.18
- [23] N. Davis, "Driving Quality Improvement and Reducing Technical Debt with the Definition of Done," 2013 Agile Conference, Nashville, TN, 2013, pp. 164-168, DOI: 10.1109/AGILE.2013.21
- [24] A. Poth, A. Sunyaev, "Effective Quality Management: Value- and Risk-Based Software Quality Management," in IEEE Software, vol. 31, no. 6, pp. 79-85, Nov.-Dec. 2014. DOI: 10.1109/MS.2013.138
- [25] G. A. Moore, "Crossing the Chasm: Marketing and Selling Disruptive Products to Mainstream Customers", Harper Business, 3rd Edition, 2014, ISBN 978-0062353948
- [26] E. Ries, "The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses," Currency, 2017, ISBN 978-1524762407
- [27] K. Ely, "Evaluating leadership coaching: A review and integrated framework," In: The Leadership Quarterly. 21, 2010. doi:10.1016/j.leaqua.2010.06.003
- [28] S. W. Ambler, "Scaling agile software development through lean governance," 2009 ICSE Workshop on Software Development Governance, Vancouver, BC, 2009, pp. 1-2. DOI: 10.1109/SDG.2009.5071328
- [29] D. Talby, Y. Dubinsky, "Governance of an agile software project," 2009 ICSE Workshop on Software Development Governance, Vancouver, BC, 2009, pp. 40-45. DOI: 10.1109/SDG.2009.5071336
- [30] R. Bavani, "Governance Patterns in Global Software Engineering: Best Practices and Lessons Learned," IEEE 6. International Conference on Global Software Engineering, 2011, pp. 50-54. DOI: 10.1109/ICGSE.2011.17
- [31] S. V. Shrivastava, U. Rathod, "A risk management framework for distributed agile projects," Information and Software Technology, Volume 85 (2017), Pages 1-15, DOI: 10.1016/j.infsof.2016.12.005
- [32] T. J. Gandomani, M. Z. Nafchi, "An empirically-developed framework for Agile transition and adoption: A Grounded Theory approach," Journal of Systems and Software, Volume 107, (2015), Pages 204-219, DOI: 10.1016/j.jss.2015.06.006
- [33] R. Chen, R. Ravichandar, D. Proctor, "Managing the transition to the new agile business and product development model: Lessons from Cisco Systems," Business Horizons, Volume 59, Issue 6 (2016), Pages 635-644, DOI: 10.1016/j.bushor.2016.06.005
- [34] K. Dikert, M. Paasivaara, C. Lassenius, "Challenges and success factors for large-scale agile transformations: A systematic literature review," Journal of Systems and Software, Volume 119, (2016), Pages 87-108, DOI: 10.1016/j.jss.2016.06.013
- [35] M. Kersten, "What Flows through a Software Value Stream?," in IEEE Software, vol. 35, no. 4, pp. 8-11, July/August 2018.
- [36] W. B. Rouse, "A theory of enterprise transformation," 2005 IEEE International Conference on Systems, Man and Cybernetics, Waikoloa, HI, 2005, pp. 966-972 Vol. 1. DOI: 10.1002/sys.20035
- [37] A. Poth, "Effectivity and economical aspects for agile quality assurance in large enterprises," Journal for Software Process: Improvement and Practice, Volume 28 Issue 11, p 1000-1004, Wiley, 2016, DOI: 0.1002/smr.1823
- [38] A. Poth, M. Kottke, "How to Assure Agile Method and Process Alignment in an Organization?" Communications in Computer and Information Science, vol 896, Springer, 2018, DOI: 10.1007/978-3-319-97925-0\_35
- [39] Sutherland et al. <http://www.agilemanifesto.org>
- [40] J. Webster, R. T. Watson, "Analyzing the past to prepare for the future: Writing a literature review," MIS Quarterly, 2002, 26(2):13-23
- [41] <https://www.agileworld.de/keynotes> (better reference will included for final paper version – conference is in the future)
- [42] <https://www.ssa-e16.com/download-the-ssa-e18-soc-reporting-guide/>
- [43] <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN>

- [44] M. E. Conway, "How do Committees Invent?", *Datamation*, 14 (5): 28-31, 1968
- [45] F. T. Taylor, "The Principles of Scientific Management," New York, NY, USA and London, UK: Harper & Brothers, 1911
- [46] P. Hines, R. Nick, "The seven value stream mapping tools," *International journal of operations & production management* 17.1 (1997): 46-64.
- [47] M. Poppendieck, T. Poppendieck, "Lean Software Development: An Agile Toolkit: An Agile Toolkit," Addison-Wesley, 2003, ISBN 978-0321150783
- [48] S. Ambler, M. Lines, "Disciplined Agile Delivery: A Practitioner's Guide to Agile Software Delivery in the Enterprise," IBM Press, 2012, ISBN 978-0132810135.
- [49] <http://www.AgileModeling.com>
- [50] <https://canvanizer.com/new/lean-canvas>
- [51] N. Asthana, T. Chefalas, A. Karve, A. Segal, M. Dubey, S. Zeng, "A Declarative Approach for Service Enablement on Hybrid Cloud Orchestration Engines," *Proceedings Network Operations and Management Symposium*. IEEE 2018, DOI: 10.1109/NOMS.2018.8406175
- [52] R. Sirkiä, M. Laanti, "Adaptive Finance and Control: Combining Lean, Agile, and Beyond Budgeting for Financial and Organizational Flexibility," *48th Hawaii International Conference on System Sciences*. IEEE, 2015, DOI: 10.1109/HICSS.2015.596

# Participating in an Industry Based Social Service Program: a Report of Student Perception of What They Learn and What They Need

Miguel Ehécatl Morales-Trujillo

University of Canterbury,  
Computer Science and Software Engineering Department  
Christchurch, New Zealand  
miguel.morales@canterbury.ac.nz

Gabriel Alberto García-Mireles

Universidad de Sonora  
Departamento de Matemáticas  
Hermosillo, Mexico  
mireles@mat.uson.mx

**Abstract**—Skills demanded by the IT industry from graduates should be aligned with the curricula of Computer Science undergraduate programs. It is well-known that theoretical knowledge undergraduate students acquire during their studies needs to be complemented with practical experience; therefore, participating in university supported real life projects is a viable option for the students to get prepared for the industry. This paper reports findings from a survey applied to students who had been involved in an industry-based program meant to fulfill their graduation requirements, including the opportunity to develop a capstone project. We gathered their perceptions regarding what they learned during their studies, what they acquired in the industry-based program and what they consider useful for their current jobs. The results show that most topics are aligned between the Bachelor's degree program and the industry needs, but there is a strong separation in the cognitive levels students achieve at each stage. The paper provides insight into the needs of Computer Science students and contributes to finding ways of increasing undergraduate student satisfaction with skills acquired at university and their application in real contexts.

## I. INTRODUCTION

**A**N accelerated evolution of IT technologies demands software developers ready to get incorporated to IT workforce. The Bureau of Labor statistics of the United States of America estimates a 24% increase in demand for software developers in the period from 2016 to 2026, which is much faster than average growth [26]. Similarly, in other countries around the world the demand for software developers is growing in similar proportions. To obtain a desirable employment, a competent software developer is required to possess a wide variety of skills, such as managerial, engineering, team working and communication [1].

In undergraduate computer science (CS) and software engineering (SE) programs, educators provide experiences to support an adequate development of knowledge and skills in students. However, a crucial question remains open: how to educate software engineers to do their jobs efficiently and properly [25]. Current curriculum guidelines recommend that educational programs provide effective learning of SE skills and concepts by incorporating real-world elements, which

could be done through capstone projects and student work experience among others [1].

In capstone projects, students focus their effort on completing a significant real project while they practice learned knowledge and skills [1]. Capstone projects represent an important learning activity in SE, given that most of the SE concepts are abstract in nature. In consequence, they are difficult to understand by undergraduate students without adequate hands-on experience [17].

However, providing a real-life experience in undergraduate programs is a difficult task [9]. Another factor to be considered is the gap between skills acquired by IT graduates and skills demanded by the industry [25]. Although it is necessary to prepare students for incorporating into the software industry, this is still a very complex endeavor that many universities are struggling with [38]. Other than theoretical knowledge, future employers are always more interested in students who are equipped with hands-on experience [41].

Despite a considerable interest to find ways of helping students to quickly become efficient software developers and blend into the IT work place, there are few empirical reports of problems that students face in capstone projects as well as of learning outcomes students perceive to have achieved after completing a capstone project [37]. Thus, this paper aims to present an experience report of how a joint effort between industry and university can increase undergraduate student satisfaction with skills learned at university and their application in a real project.

In addition, this report provides a comparison between knowledge and skills learned at university against those learned in the industry. Moreover, the results of this experience have other effects over educational aspects, such as an increase in awareness for alignment of academic courses with industrial experience by CS and SE educators.

The paper is structured as follows. Section II presents an overview of capstone projects and agile methodologies used in them. Section III describes the context of CS Bachelor's degree of a Mexican university and the nature of the industry-based student program. Section IV reports the data collection process

and data analysis. The results are discussed in Section V and the conclusions and future work are presented in Section VI.

## II. BACKGROUND

### A. Capstone projects in CS

Capstone is defined as “the high point: crowning achievement” [24]. In a capstone project, “students solve real-life problems in the context of a large, realistic project” [37]. Working on a project helps students to develop SE skills and apply them in a realistic environment, providing them with an excellent opportunity to understand and experience various technological waves that would be present in their careers [16]. In order to improve learning outcomes, project courses can involve industry partners who provide real-world problems for students to solve, to develop both technical and soft skills, and to gain contacts to potential mentors in the industry [40].

Several papers address the use of capstone projects to enhance knowledge and skills of both CS and SE students. Vanhanen et al. [37] conducted a survey to understand the problems that arise during development of capstone projects, to gain insight into improvement in learning outcomes as perceived by students, and into customer satisfaction. They found that learning outcomes varied a lot among teams and team members since roles undertaken by students affected their level of learning. Considering common problems reported in capstone projects, the most common ones were related to testing, task management, effort estimation and technology skills [37].

In [21] students identified algorithms, programming, networks and databases, and contemporary technologies for developing software as essential technical skills in order to succeed in a capstone project. They can be seen as potential discipline knowledge gaps and become areas to better address in capstone project activities. The authors surveyed students’ feedback where the majority of students reported a general skill improvement or learning new technical or project management skills during their capstone project [21].

Projects often require students to use technologies (programming languages, web frameworks etc.) that they have not been taught in their previous courses [37]. Based on observing differential learning outcomes achieved by students, Karunasekera and Bedse [17] proposed a skill based learning framework to provide objective guidance to ensure that team based projects offer a balance of management, engineering and personal skills. In case of capstone project courses, Majanoja and Vasankari [21] suggest to organize them focusing on team work, communication and problem solving skills.

It can be observed that few quantitative empirical studies report problems encountered by capstone project teams [37]. The reported problems are related to poor communication among the team members, poor leadership, failure to compromise, procrastination problems, integration testing problems, and lack of cooperation, among others [37]. Besides, such problems as lack of skills for using tools, lack of organization, lack of technology expertise, and having to combine work and

study during the project are reported in [2]. Other risks are architecture complexity, quality trade-offs, personnel shortfalls, budget and schedule constraints, COTS and other independently evolving systems, customer-developer-user team cohesion, requirements volatility, user interface mismatch, process quality assurance, requirements mismatches, acquisition and contracting process mismatches [18].

Learning outcomes of capstone projects [37] include improvement in such students skills as familiarity with agile approach, programming, project planning and management, effort estimation, acquiring “a big picture” of a software development process, team work, customer interaction, and communication [20]. Improvement in learning is reported in skills related to requirements engineering, system design, modeling, programming, version control, release management, and usability engineering, among others [6]. Besides, Broman et al. report students’ learning technical knowledge of a specific SE role, time management, usefulness of agile methods, and team communication and collaboration [5]. As for guidelines to discuss student experience of participation in industrial capstone projects, SWEBoK offers a set of topics, understandable to students and widely used by researchers [32].

A realistic setting of a capstone project developed for a real customer also introduces challenges for the participants. The beginning of a project course is particularly difficult, as students have to familiarize themselves with development processes and tools, get to know the project’s problem domain, understand requirements, and deal with communication issues [9]. Besides, there are several organizational aspects to be considered when providing real-life experiences [38], [6], [4]. Despite the effort to provide students with real-world experiences, instructors should be aware that the level of exposure to these formative actions will be limited [1]. Instructors can only facilitate the initial stage of the process for students develop a mature understanding of the real world across their careers [1].

### B. Agile methodologies in capstone projects and courses

Providing students with an iterative work by means of agile and lean methods can both fulfill the industry needs and support the design of SE courses [23]. Considering agile methodologies, Scrum is one of the most adopted and/or adapted in a capstone course design [23], [27]. Indeed, in courses based on project-based learning, Scrum is very common. It can provide control over project progress and it is able to ensure a steady pace of development [12].

Fitsilis and Lekatos [13] conducted a survey study to understand the importance of agile practices and their relevance in the industrial context while relying on the Scrum method taught in a capstone course. As a result, the majority of the participants reported a general positive experience from the capstone course; in addition, they perceived such frequently used in the industry practices as unit testing, coding standards, test-driven development (TDD) and continuous deployment as the most useful. Other researchers studied how particular estimation techniques can be introduced in a software engineering

capstone course [28]. Another research line is focused on the assessment of particular practices of continuous integration, TDD, and work tracking, among others [12].

Both, supporting tools and adaptation of agile techniques, have been explored in capstone courses and projects. Rodríguez et al. [29] propose a virtual Scrum tool that simulates a Scrum-based team room by means of the 3D virtual world metaphor. The tool can support specific configuration and progress of each student team [29]. On the other hand, Krusche et al. [19] proposed Rugby, an agile model for supporting continuous delivery of software. The authors report that the Rugby method improves coordination across multiple teams and enhances communication between developers and customers [19].

Lean approaches, such as Kanban, have also been studied in the context of Scrum-based university courses by introducing Kanban at the end of instruction period [23]. Paasivaara et al. [27], on the other hand, report an effective outcome of a Scrum based capstone course with real clients. They found that students positively change their attitudes about the importance of collaboration and communication within the teams while experiencing less than expected difficulty in learning new technologies [27].

Moreover, agility in capstone courses have been explored in relation to other topics. Fagerholm and Vihavainen [11] have studied peer assessment to provide a full view of student performance in a project. In particular, self-assessment and peer review should provide insights with regard to learning goals. In software maintenance, Weissberger et al. [39] reported an experience of using agile principles in a maintenance project developed in a capstone project.

Academic software factory is another approach used in universities to enhance the quality of education by means of capstone projects. University laboratories emulate a real working setting where teams of students implement a project for real customers and real deadlines [33]. Similarly, Fagerholm et al. [10] report another experience of developing the software factory approach. They provide a collection of patterns and anti-patterns to support the design, implementation and operation of project-based startups [10].

### III. CONTEXT OF THE EXPERIENCE

#### A. Computer Science Bachelor's degree

The National Autonomous University of Mexico (UNAM) is the biggest university in the country with 356,530 enrolled students in 2019 [35] and sits the #2 in Latin America (#113 university in the world) according to QS World University rankings [7]. The UNAM consists of 15 colleges that offer 127 Bachelor's degrees; CS Bachelor's degree is one of them and is taught at the College of Science.

CS Bachelor's degree is designed to be finished in four years obtaining 376 credits; it consists of 28 compulsory and 6 elective courses based on seven knowledge areas: Mathematical foundations, Discrete structures, Programming, Software engineering, Theoretical computing, Theory and practice integration, and Systems architecture. In order to graduate, CS

students are required to complete 376 credits, 480 hours of social service and to develop a capstone project.

In Mexico, undergraduate students must meet the requirement of completing a social service to be able to graduate [8]. The goal of the social service program is to contribute to both academic and professional training of students. Student who have achieved 70% of undergraduate credits are eligible for social service program as they are supposed to apply their acquired knowledge and skills in an organization which is looking for social, cultural and economic development. The social service practice amounts to 480 hours that students complete in a time period of 6 to 24 months [34].

Mexican program of social service "promotes professional and human development of the student, developing an active and supportive social commitment applied to the solution of problems or needs of the country" [34]. It is similar to work experience placements required by universities around the world, however it differs from them due to its civic and social orientation. Nevertheless, there is no other alternative that would provide students with work experience previous to their graduation.

As for the capstone project, CS students are offered 10 options: writing a dissertation (thesis), participating in a research-based project, taking a research seminar, participating in a science popularization project, reaching the national stage of the ACM Programming Contest, being in the first quartile of the Graduate Record Examination for CS, obtaining an over 95 GPA (out of 100), carrying out tertiary teaching support activities (such as teaching assistant), obtaining industry experience and extending the social service.

The majority of these options are research- or teaching-based with the exception of the last two: industry experience and extended social service. In the case of the extended social service, most of the programs are research and teaching oriented. According to the information published in [36], 97 out of 272 available programs for CS students are IT related. From those, 80 are carried out within the UNAM schools and dependencies, 16 in government offices and only one is industry-based.

In addition, both options require a university professor who can supervise capstone projects based on an industry experience or an extended social service, however, there is not enough cooperation between university supervisors and industry available projects.

We identified a lack of opportunities for students to gain more industry-oriented experience both through social service or developing a capstone project. In order to improve this situation, we proposed to push for industry-based capstone projects and social service programs.

On the one hand, it would allow students to complete a social and civic service requirement as well as to obtain work experience in the current industry. On the other hand, it can lead to an increase in numbers of graduated students in CS as industry-based experience offers reasonable opportunities for developing a successful capstone project.

The current numbers of graduated students are extremely low: the terminal efficiency of CS Bachelor's degree since 1995 is 18.1%. In absolute numbers we find that 285 students graduated, 621 completed their credits but did not graduate and 672 neither graduated nor completed their credits. Only one of each five students is able to complete their credits, get through their social service and develop a capstone project. Looking at only eight last generations of students, the terminal efficiency is slightly higher: 19.7%. while only 155 out of 785 CS students graduated.

One of the reasons for the low numbers of IT-related graduates is their high level of employability, and CS students are no exception. Nowadays, there is an important demand to fill in constantly growing job offers in the IT sector. Deficit of employees with a IT profile is critical in many countries around the world; according to [25] it is growing at an exponential rate. Students are leaving the university because they are getting job offers even if they are not graduated yet.

In this context, undertaking this initiative was powered by a strong motivation to improve the exposure of CS students to relevant work experience in real projects aligning it with their graduation requirements, namely, social service and capstone project.

### B. Industry based program

The industry based student program described in this paper was divided in two stages. During the first stage students worked to complete 480 hours of social service. During the second stage students could opt for continuing working but with a condition of developing a capstone project based on the work done in the first stage.

This project was carried out in a joint effort by the College of Science and a Mexican software development organization. The expected outcome of the project was an information system for public administration. The first step to link the organization and the university was to define a social service program in which the student activities would be aligned with their civic responsibilities.

Once the program was created and approved by the University, any interested student could join the program. Actually, students joined the program voluntarily at different stages of the project during the year. There were no additional restrictions to join the program out of those already imposed by the University, so all students who applied for the program were accepted.

Each student was coordinated by a professor from the College and looked for the goals of the social service to be achieved. Students were required to work in the organization during a minimum of 6 months covering the 480 hours of their social service. They also received a monthly grant during the time they participated in the project.

The organization had a full development process in place; in consequence, the students were able to carry out a wide variety of tasks in different software engineering activities. During their first month of work, they were trained and introduced into the organization, the team and the project. Tasks assigned

during the initial period were related to requirements specification and testing with the intention to get them familiar with the project and system under development. Activities related to Software Configuration Management (SCM), such as version control systems and continuous integration, were specially relevant at this stage.

The organizational development process was a hybrid process based on [22] mixing agile practices with ISO/IEC 29110-5-1-2. Project management activities were assigned using boards (Kanban style) and the progress was reported through stand-ups (Scrum). Software requirements specification (SRS) was carried out first through wireframes and customer workshops and then specified as use cases. Programming tasks consisted in fixing bugs and pair programming sessions arranged on demand while continuous integration and deployment practices were in place. User acceptance tests were carried out mostly by using a Think Aloud! protocol, bughunt sessions to test the system were implemented, and a bug tracker system was used to report defects.

During the following months students rotated between different teams and carried out roles of analysts, programmers, testers and database developers. Table I enlists the tasks carried out by the students, where agile practices used for each task are provided in brackets. For standardizing purposes generic ISO/IEC 29110-5-1-2 [14] tasks definitions are used.

A student's timeline in the program was usually the following:

- **Day 1:** an introduction to the organization, team and project.
- **Day 2:** a walk-through the development process of the organization.
- **Day 3:** joining SRS and Testing teams.
- **Month 2:** joining Construction team.
- **Month 3:** joining Databases team.
- **Months 4 and 5:** joining the team of their preference.
- **Last two weeks:** writing a social service report required by the College of Science.

Students received training from the team leaders and participated in workshops obtaining knowledge in topics like:

- Version control systems with Git and continuous integration.
- Development frameworks with Grails.
- Web deployment using WebLogic Server.
- Stored procedures and packages in Oracle 12c.
- Test automation with JMeter.
- Mobile development.
- Basic accounting to deal with employment and workers' rights.

The students had regular control meetings with the project manager of the organization following a daily Scrum approach. In addition, they had regular contact with their respective team leaders and teammates.

From the College of Science side, two professors were in charge of ensuring that the purpose of the social service program was being fulfilled and the students received guidance

TABLE I  
TASKS CARRIED OUT BY STUDENTS

<b>Software Requirements</b>
Document or update requirements specification ( <b>User stories and Wireframes</b> )
Identify and consult information sources in order to get new requirements.
Verify and obtain approval of the requirements specification.
Validate that requirements specification satisfies needs and agreed upon expectations, including the user interface usability ( <b>Think Aloud!</b> ).
Participate in revision meetings with the customer.
<b>Software Architectural and Detailed Design</b>
Describe in detail the appearance and the behaviour of the UI ( <b>Wireframes and Think Aloud!</b> ).
Generate an architectural design, its arrangement in subsystems and components defining internal and external interfaces.
<b>Software Construction</b>
Construct or update software components ( <b>CI and CD</b> ).
Correct the defects found until successful unit test is achieved ( <b>Pair programming</b> ).
Perform backup according to the version control strategy.
<b>Software Integration and Tests</b>
Verify consistency among requirements specification, software design and test cases ( <b>Think Aloud!</b> ).
Design or update unit test cases.
Perform software tests and document results in the test report ( <b>Bughunt</b> ).
Correct defects found and perform regression tests ( <b>Pair programming</b> ).
Verify consistency of the software documentation with the software.
<b>Product Delivery</b>
Verify consistency of maintenance documentation with software configuration.

from the organization members. A non-conformity process was in place to inform any improvement opportunity or disagreement regarding the students' interaction and performance. This process was a two way communication channel that served to inform the College if students did not perform as expected or if they were absent without notifying.

Once the six months concluded and students completed the required 480 hours of their social service, they were invited to continue in the organization but developing their capstone project. The capstone project had to be related to their work in the last six months and was supervised by a professor from the College with a possibility to be co-supervised by a member of the organization. The only constraint imposed on students was to finish the capstone project in no more than six months.

This paper reports the participation of a cohort formed by 13 students. The time they spent in the program varied from 5 to 12 months. First results of the program are described in the following sections.

#### IV. MAPPING ACQUIRED AND REQUIRED SKILLS

As mentioned before, there is a need for aligning higher education practices to industry needs by exposing students to industrial processes [30]. It is often observed that a large gap separates software projects in industry from what can be experienced in the classroom [15]; therefore, students are unconvinced by the relevance of the material delivered in lectures. The challenge, in consequence, is to develop environments within universities that are sufficiently real to be convincing to students [31].

Our interest was to gain an insight into students' perceptions regarding the knowledge and skills they acquired during their

studies, during their participation in the program, and finally in their current job. To gather data, we designed a survey based on SWEBoK areas and topics; we were interested to know to what extent the students have been exposed to them, how the students perceived the preparation they received and what they considered necessary once they joined the industry.

The results were expected to allow for evaluation of the usefulness of the social service program as well as for an appreciation of an (non)-existing alignment between what CS students perceive that have learned and what they need once they graduated.

SWEBoK establishes a baseline for the body of knowledge in the field of SE [32]. It strongly influences the manner in which curricula are defined and how academic programs are accredited. In this study, SWEBoK was chosen due to its wide acceptance as a well-known SE body of knowledge that connects industry recommended practices and the knowledge expected from a software engineering professional.

The survey consisted of two sections, the first section was an online survey and the second section was answered in a spreadsheet. The time taken to answer the whole survey ranged from one day to one week; an incentive was given for each fully completed survey.

The first section contained questions about personal involvement in the program, and their current academic and professional situation:

- 1) How many months were you involved in the program?
- 2) What activities did you carry out during your participation in the program?
- 3) What skills do you think you developed the most?

- 4) Did you develop a capstone project derived from your participation in the program?
- 5) Did you complete the social service program?
- 6) Did this program contribute or is contributing to your graduation?
- 7) Did this program contribute or is contributing to you getting a job?
- 8) Do you currently work in the IT-related industry?
- 9) What do you consider to be the most beneficial in your participation in the program?
- 10) What improvements do you consider the program needs?

In the second part, the students were presented with a table which enlisted each of the SWEBoK topics grouped by area of knowledge, and had three columns representing stages (while studying my bachelor's, while participating in the program, in my job), during which the knowledge was acquired and/or applied.

Each SWEBoK topic (rows) was presented with an example in the form of tooltips. Figure 1 shows a fragment of the survey with answers retrieved from one student. The fragment was translated from Spanish into English.

Once the collected data were analyzed through general descriptive statistics of median, mode and mean, they were used to answer the following research question:

*How did participating in a real project transform the students' knowledge?*

It was decided to follow an adapted Bloom's taxonomy presented in [3], particularly three lower levels of the taxonomy (Remember, Understand and Apply) were used.

In the first column, respondents selected one of the following options for each SWEBoK topic to complete the claim "While studying my bachelor's ...":

- 1) I did not get familiar with this topic.
- 2) I got familiar with this topic but I did not apply it in practice (**Remember and Understand**).
- 3) I got familiar with this topic and had a chance to apply it in practice (**Apply**).

In the second column, the options to select were similar to the one in the previous question, but the claim was "While participating in the program ...":

- 1) I did not require this topic.
- 2) I was already familiar with the topic, but I applied it in practice (**Apply**).
- 3) I was already familiar with the topic, but I did not apply it in practice (**Remember and Understand**).
- 4) I got familiar with the topic and I had the chance to apply it in practice (**Remember, Understand and Apply**).

The second question of interest was:

*What software engineering topics/areas are regarded as most useful in the industry by the students?*

In this case, a 5-point Likert scale was used. In the third column, respondents selected one of the following options to complete claims corresponding to "In my job, <SWEBoK topic> is ...":

- 1) Not useful.

- 2) Slightly useful.
- 3) Useful.
- 4) Very useful.
- 5) Essential.

Besides it was of interest to know what positions they held in the industry once they finished their participation in the project. This information was collected via email.

## V. RESULTS AND DISCUSSION

### A. First section of the survey: Industry-based program

A total of 13 students participated in the program and 12 completed it successfully; Table II shows a summary of the students who participated in the program. The only student who did not complete the program accepted a job offer in the IT industry and left the program. This is one of main reasons of why students do not graduate and there is very little the universities can do to minimize its impact.

The first part of the survey was responded by 10 students. Eight out of 10 students reported that their participation in the program contributed to their graduation and nine out of 10 reported that it contributed to getting a job after the program. The following statements are examples of the students' opinions regarding their experience in the program:

*"Having a real-world experience that goes beyond a social service gave me a better position."*

*"The experience I have got was well valued by the recruiter."*

*"I met people with a lot experience in software development, I learned a lot from them."*

*"I interacted with real customers and teams, the experience I have got was real-life."*

Six students provided information regarding the position they occupied after the project:

- Database Developer
- SOA Java Developer
- Software Test Engineer
- Data Scientist Jr.
- Software Engineer
- SECaaS and DBaaS Engineer

Two of them worked in an organization that followed Scrum as its main methodology, allowing one of them to become a Scrum Master.

Last but not the least, four students concluded their capstone projects as based on their work during the program. Given the proportion of students who completed their social service (92.3%) and those who graduated (30.8%) we consider this an improvement for the historical data of graduated CS bachelors in the UNAM.

### B. Second section of the survey: SWEBoK areas and topics

Nine students completed the SWEBoK mapping. The first question to answer is:

*How did participating in a real project transform the students' knowledge?*

On a more general level, Table III displays an association perceived by students between each SWEBoK area

	0: I did not get familiar with this topic. 1: I got familiar with this topic but I did not apply it in practice. 2: I got familiar with this topic and had a chance to apply it in practice.	0: I did not require this topic. 1: I was already familiar with the topic, but I applied it in practice. 2: I was already familiar with the topic, but I did not apply it in practice. 3: I got familiar with the topic and I had the chance to apply it in practice.	1: Not useful. 2: Slightly useful. 3: Useful. 4: Very useful. 5: Essential.
<b>1. Software Requirements</b>	<b>While studying my bachelor's ...</b>	<b>While participating in the program ...</b>	<b>In my job ...</b>
1.1. Software Requirements Fundamentals	2	1	3
1.2. Requirements Process	2	1	3
1.3. Requirements Elicitation	1	3	5
1.4. Requirements Analysis	2	1	2
1.5. Requirements Specification	2	1	4
1.6. Requirements Validation	1	3	5
1.7. Practical Considerations	1	3	4
1.8. Software Requirements Tools	2	1	5

Fig. 1. A survey fragment

TABLE II  
PARTICIPANTS' DETAILS

ID	# Months	Roles carried out	Completed the program?	Developed a capstone project?	Got employed?
S1	5	Analyst and DB developer	No	No	Yes
S2	12	Analyst, Programmer and Tester	Yes	Yes	Yes
S3	8	Analyst and Programmer	Yes	Yes	Yes
S4	6	Analyst and Tester	Yes	No	Yes
S5	6	Analyst, Tester and DB developer	Yes	No	Yes
S6	7	Analyst and Tester	Yes	No	Yes
S7	6	Analyst and Tester	Yes	No	Yes
S8	7	Analyst and DB developer	Yes	No	Yes
S9	9	Analyst, Programmer and Tester	Yes	No	Yes
S10	11	Analyst, Programmer and Tester	Yes	Yes	Yes
S11	10	Analyst, Programmer and Tester	Yes	Yes	Yes
S12	9	Analyst and Tester	Yes	No	Yes
S13	6	Analyst and Tester	Yes	No	Yes

and cognitive levels. It is observed that, except for Software Maintenance, all the areas required in the program had been taught at the BSc (students got familiar with them) and most of them reached the next cognitive level (students applied them in practice).

It can be noted that Software Maintenance was perceived as an area not familiar to students during the BSc; however, by the end of the program, students perceived a shift from not knowing it to applying it in practice. Software Engineering Economics is another area perceived as unknown by students, and this area of knowledge was not required during the program either.

Students reported that after participating in the program, cognitive levels improved in nine areas and remained at the same level in three. These results allow to conclude that, in this particular case, real experience in a controlled environment offered an important advantage for the student development.

Going down on a more detailed level, we found an increase in cognitive levels associated to SWEBoK topics as 45 topics were perceived as improved and for 25 their cognitive levels remained the same. Only topics required during the project were analyzed (70 out of 102 SWEBoK topics).

In particular, three topics changed their states from Unknown to Apply, as well as from Unknown to Remember

and Understand. Their breakdown is shown in Table IV. An important message is the lack of Software Maintenance and Software Configuration Management in the curricula. The gap between students' knowledge acquired during their studies is large as compared to the importance of this area for their job in the industry. This could be caused by the lack of opportunities to work in long projects and over an existent software product, where it is possible to show the effects of good/bad maintenance practices or SCM strategies.

The second question to answer was:

*What software engineering topics/areas are regarded as most useful in the industry by the students?*

The data showed that Software Engineering Professional Practice is the only area with a median of 5 (essential), while eight other areas obtained a 4 (very useful). It is worth mentioning that the least useful, according to the student perception, turned out to be Engineering Foundations. The analysis consisted in calculating the median of topics per each area; see Table V for the results.

A more granular analysis performed on the topic level showed that nine topics got a median of five (see Table VI).

Students reported to know all of those topics from their university background, however, only when participating in the project they had an opportunity to apply three of them in

TABLE III  
COGNITIVE LEVELS OF SWEBoK AREAS ACCORDING TO STUDENT PERCEPTIONS

SWEBoK area	Cognitive level at the end of BSc	Cognitive level at the end the program
1: Software Requirements	1. Remember and 2. Understand	3. Apply
2: Software Design	1. Remember and 2. Understand	3. Apply
3: Software Construction	3. Apply	3. Apply
4: Software Testing	1. Remember and 2. Understand	3. Apply
5: Software Maintenance	0. Unknown	3. Apply
6: Software Configuration Management	1. Remember and 2. Understand	3. Apply
7: Software Engineering Management	1. Remember and 2. Understand	3. Apply
8: Software Engineering Process	1. Remember and 2. Understand	3. Apply
9: Software Engineering Models and Methods	3. Apply	3. Apply
10: Software Quality	1. Remember and 2. Understand	3. Apply
11: Software Engineering Professional Practice	1. Remember and 2. Understand	3. Apply
12: Software Engineering Economics	0. Unknown	0. Not required
13: Computing Foundations	3. Apply	3. Apply
14: Mathematical Foundations	3. Apply	0. Not required
15: Engineering Foundations	1. Remember and 2. Understand	0. Not required

TABLE IV  
SWEBoK TOPICS WITH IMPROVED COGNITIVE LEVELS

SWEBoK topics
<b>Unknown</b> $\rightarrow$ <b>Remember and Understand</b>
10.4. Software Quality Tools
<b>Unknown</b> $\rightarrow$ <b>Apply</b>
5.5. Software Maintenance Tools
6.1. Management of the SCM Process
6.5. Software Configuration Auditing

TABLE V  
SWEBoK AREAS SORTED BY USEFULNESS IN THE INDUSTRY

SWEBoK area	Median
11: Software Engineering Professional Practice	5
1: Software Requirements	4
2: Software Design	4
3: Software Construction	4
4: Software Testing	4
6: Software Configuration Management	4
7: Software Engineering Management	4
8: Software Engineering Process	4
13: Computing Foundations	4
5: Software Maintenance	3
9: Software Engineering Models and Methods	3
10: Software Quality	3
14: Mathematical Foundations	3
12: Software Engineering Economics	2
15: Engineering Foundations	1

TABLE VI  
SWEBoK TOPICS CONSIDERED TO BE ESSENTIAL IN THE WORK PLACE

SWEBoK topics
3.3. Practical Considerations
3.4. Construction Technologies
6.6. Software Release Management and Delivery
11.2 Group Dynamics and Psychology
11.3. Communication Skills
13.3. Programming Fundamentals
13.4. Programming Language Basics
13.12. Database Basics and Data Management
14.2. Basic Logic

TABLE VII  
SWEBoK TOPICS THAT STUDENTS KNEW BUT HAD NEVER APPLIED BEFORE THE PROGRAM

SWEBoK topics
6.6 Software Release Management and Delivery
11.2 Group Dynamics and Psychology
11.3 Communication Skills

practice (see Table VII).

In the context of software development, Software Release Management and Delivery constitutes a fundamental area of knowledge for every practitioner. On the other hand, Group Dynamics and Psychology together with Communication Skills are abilities strongly required for teamwork. It was a favorable outcome that students were presented with an opportunity to applied these topics during the project.

The most developed during the project areas are shown in Table VIII while the least developed ones are presented in Table IX.

TABLE VIII  
SWEBoK AREAS THAT STUDENTS DEVELOPED THE MOST

SWEBoK topics
1: Software Requirements
2: Software Design
3: Software Construction
4: Software Testing
5: Software Maintenance
6: Software Configuration Management
7: Software Engineering Management
8: Software Engineering Process
9: Software Engineering Models and Methods
10: Software Quality
11: Software Engineering Professional Practice
13: Computing Foundations

TABLE IX  
SWEBoK AREAS THAT STUDENTS DEVELOPED THE LEAST

SWEBoK topics
12: Software engineering economics
14: Mathematical foundations
15: Engineering foundations

### C. Limitations and Threats to Validity

Construct validity: the data collection, particularly the survey on SWEBoK areas and topics, included examples of each topic presented to students for clarification. Also, in order to properly define the cognitive level of each SWEBoK topic, an adaptation of the Bloom's taxonomy was used. Although the insights of this experience are based on a limited number of sources, the data were obtained directly from the main stakeholders, namely the CS students. However, the results, in order to be generalized, require more generations of students to join similar programs.

External validity: it is worth to mention that the social service is specific for the Mexican context, however, it is representative of the country. Besides, we consider this program to be an initial step towards demonstrating the usefulness of running industry-based programs within university contexts.

Internal validity: the students joined the program voluntarily, and the information about the program was disseminated in the same way as the rest of the programs. A distinctive feature of this program was the grant offered to the students, which is not common to the majority of the programs; therefore, it could be a factor for the students to choose the program.

The findings are based on the perception of students and could be improved by applying specific assessments at certain phases of the program. Nevertheless, an important advantage of this study relies on the fact that we traced student data from studying a bachelor's degree till joining the IT workforce, which provides a deeper insight into the problem.

## VI. CONCLUSIONS

In Mexico, university graduation requirements consist of developing a capstone project in the form of a thesis and its oral defense, completing credits of the BSc and 480 hours of social service.

As an alternative for providing students with relevant working experience, the UNAM College of Science put in place a social service program run in conjunction with a software development organization. In this program students had the opportunity to cover their graduation requirements while being immerse in a real project. In this first experience, 13 students participated representing around 11% of the CS students of a generation (114 students).

A total of 12 students completed their social service, which was the primary goal of the program, and 4 students finished their capstone projects and graduated (30.8% of the students who joined the program).

Despite differences across countries, there is definitely a growing interest in establishing a university-industry collaboration in order to promote well-prepared graduates, where these initiatives are welcomed by the students [30]. Teachers in charge of capstone project courses could benefit from a better understanding of what kind of problems students typically encounter in capstone projects [37].

We hope that this type of experiences will help to increase academic success, improve students' skills, reducing numbers of dropouts. It is worth to mention that all the students who participated in the program were employed in the IT sector after finishing the program.

Finding of the study showed that area 12: Software Engineering Economics was neither required during the project nor covered by the curricula according to the participants' perception. On the other hand, two important areas are perceived as not covered by the curricula but were required during the program: 5: Software Maintenance and 6: Software Configuration Management.

Finally, there is a clear cut division between cognitive levels developed by the students: *remember* and *understand* during their studies, *apply* when working. It was an expected outcome; however, it reinforces the idea of looking for alternatives for providing students with opportunities to apply their knowledge in practice in order to obtain an integral education.

As future work it is expected to run more programs involving IT companies in order to demonstrate the usefulness of aligning social service with relevant work experience, thus helping students to achieve their academic goals and to successfully join the industry. Besides, it is expected to explore potentials of this approach in other universities in the country. In parallel, an in-depth analysis of the current curricula should be carried out with the aim of including such missing topics as Software Maintenance and Software Configuration Management.

### ACKNOWLEDGMENT

The authors would like to thank Mauricio Morgado and Ricardo Cruz for leading the project, the professors Guadalupe

Ibargüengoitia and Hanna Oktaba for coordinating the capstone projects. Also, thanks to the CS students that participated in this initiative for their contribution: Gerardo González, Adolfo Marín, Aarón Guerrero, Alan Gutiérrez, Jhovan Gallardo, Julio Chávez, Rafael Robles, Diana Góngora, Marco Estrada, Edson Servín, Edgar Tapia and Rodrigo Casiano.

## REFERENCES

- [1] ACM-IEEE. Software engineering curriculum guidelines, 2014.
- [2] Tero Ahtee and Timo Poranen. Risks in students' software projects. In *22nd Conference on Software Engineering Education and Training*, 154–157. IEEE, 2009.
- [3] Lorin Anderson, David Krathwohl, Peter Airasian, Kathleen Cruikshank, Richard Mayer, Paul Pintrich, James Rath, and Merlin Wittrock. A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Abridged edition, Longman, 2001.
- [4] Barry Boehm and Daniel Port. Educating software engineering students to manage risk. In *23rd International Conference on Software Engineering*, 591–600, IEEE, 2001.
- [5] David Broman, Kristian Sandahl, and Mohamed Abu Baker. The company approach to software engineering project courses. *IEEE Transactions on Education*, 55(4):445–452, 2012.
- [6] Bernd Brügge, Stephan Krusche, and Lukas Alperowitz. Software engineering project courses with industrial clients. *TOCE*, 15:17:1–17:31, 2015.
- [7] QS Crimson. QS World university rankings, 2019.
- [8] Diario Oficial de la Federación (Official Journal of the Federation). Reglamento para la prestación del servicio social de los estudiantes de las instituciones de educación superior en la República Mexicana (Regulation for the social service provision of the students of the tertiary education institutions in Mexico), 1981.
- [9] Dora Dzonyar and Bernd Bruegge. Reaching steady state in software engineering project courses. In *Combined Workshops of the German Software Engineering Conference (SE 2018)*, 8–11, 2018.
- [10] Fabian Fagerholm, Arto Hellas, Matti Luukkainen, Kati Kyllönen, Sezin Yaman, and Hanna Mäenpää. Designing and implementing an environment for software start-up education: Patterns and anti-patterns. *Journal of Systems and Software*, 146:1 – 13, 2018. <https://doi.org/10.1016/j.jss.2018.08.060>
- [11] Fabian Fagerholm and Arto Vihavainen. Peer assessment in experiential learning assessing tacit and explicit skills in agile software engineering capstone projects. In *2013 IEEE Frontiers in Education Conference (FIE)*, 1723–1729. IEEE, 2013. <https://doi.org/10.1109/FIE.2013.6685132>
- [12] Vinícius Gomes Ferreira and Edna Dias Canedo. Design sprint in classroom: exploring new active learning tools for project-based learning approach. *Journal of Ambient Intelligence and Humanized Computing*, 1–22, 2019. <https://doi.org/10.1007/s12652-019-01285-3>
- [13] Panos Fitsilis and Alex Lekatos. Teaching software project management using agile paradigm. In *21st Pan-Hellenic Conference on Informatics*, 47:1–47:6, ACM, 2017. <http://doi.acm.org/10.1145/3139367.3139413>
- [14] ISO/IEC TR 29110-5-1-2:2011 software engineering – lifecycle profiles for very small entities (VSEs) – part 5-1-2: Management and engineering guide: Generic profile group: Basic profile, ISO, 2011.
- [15] M.J.I.M. Genuchten, van and D.R. Vogel. Getting real in the classroom. *Computer*, 40(10):106–108, 2007.
- [16] Carlo Ghezzi and Dino Mandrioli. The challenges of software engineering education. In *27th International Conference on Software Engineering*, 637–638, ACM, 2005. <http://doi.acm.org/10.1145/1062455.1062578>
- [17] Shanika Karunasekera and Kunal Bedse. Preparing software engineering graduates for an industry career. In *20th Conference on Software Engineering Education and Training (CSEET'07)*, 97–106. IEEE, 2007.
- [18] Supannika Koolmanojwong and Barry W. Boehm. A look at software engineering risks in a team project course. *26th International Conference on Software Engineering Education and Training (CSEET)*, 21–30, 2013.
- [19] Stephan Krusche, Lukas Alperowitz, Bernd Bruegge, and Martin O Wagner. Rugby: an agile process model based on continuous delivery. *RCoSE*, 14:42–50, 2014.
- [20] Viljan Mahnic. A capstone course on agile software development using scrum. *IEEE Transactions on Education*, 55:99–106, 2012.
- [21] Anne-Maarit Majanoja and Timo Vasankari. Reflections on teaching software engineering capstone course. In *10th International Conference on Computer Supported Education*, volume 2 of *CSEDU 2018*, 68–77, 2018.
- [22] Erick Matla-Cruz, Miguel Morales-Trujillo, and David Velázquez-Portilla. Disciplinando equipos pequeños con prácticas ágiles (agile practices and small teams discipline). *Difusión*, 8(2):28–33, 2014.
- [23] Christoph Matthies. Scrum2kanban: integrating kanban and scrum in a university software engineering capstone course. In *2nd International Workshop on Software Engineering Education for Millennials*, 48–55, ACM, 2018.
- [24] Merriam-Webster.com. Capstone, 2019.
- [25] Ana M Moreno, Maria-Isabel Sanchez-Segura, Fuensanta Medina-Dominguez, and Laura Carvajal. Balancing software engineering education and industrial needs. *Journal of Systems and Software*, 85(7):1607–1620, 2012.
- [26] Bureau of Labor Statistics. Occupational Outlook Handbook, 2019.
- [27] Maria Paasivaara, Dragoş Vodă, Ville T Heikkilä, Jari Vanhanen, and Casper Lassenius. How does participating in a capstone project with industrial customers affect student attitudes? In *40th International Conference on Software Engineering: Software Engineering Education and Training*, 49–57, ACM, 2018.
- [28] Marko Požnel and Viljan Mahnič. Studying agile software estimation techniques: the design of an empirical study with students. *Global Journal of Engineering Education*, 18(2), 2016.
- [29] Guillermo Rodríguez, Alvaro Soria, and Marcelo Campo. Virtual Scrum: A teaching aid to introduce undergraduate software engineering students to Scrum. *Computer Applications in Engineering Education*, 23(1):147–156, 2015.
- [30] Manuel Rodríguez, Mario Vázquez, Hariklia Tsalapatas, Carlos de Carvalho, Triinu Jesmin, and Olivier Heidmann. Introducing lean and agile methodologies into engineering higher education: The cases of Greece, Portugal, Spain and Estonia. In *IEEE Global Engineering Education Conference*, 720–729, 2018. <https://doi.org/10.1109/EDUCON.2018.8363302>
- [31] Robbie Simpson and Tim Storer. Experimenting with realism in software engineering team projects: an experience report. In *30th Conference on Software Engineering Education and Training (CSEET)*, 87–96. IEEE, 2017.
- [32] IEEE Computer Society, Pierre Bourque, and Richard E. Fairley. *Guide to the Software Engineering Body of Knowledge (SWEBOK(R)): Version 3.0*. IEEE Computer Society Press, 3rd edition, 2014.
- [33] Davide Taibi, Valentina Lenarduzzi, Muhammad Ahmad, Kari Liukkunen, Maria Lunesu, Martina Matta, Fabian Fagerholm, Jürgen Münch, Sami Pietinen, Markku Tukiainen, Carlos Fernández, Juan Garbajosa, and Kari Systä. “Free” innovation environments: Lessons learned from the software factory initiatives. In *International Conference on Software Engineering Advances IARIA*, 2015.
- [34] DGOSE UNAM. Dirección General de Orientación y Atención Educativa (General Directorate of Educational Orientation and Attention), 2017.
- [35] DGPE UNAM. Portal de estadística universitaria (Statistics web portal of the University), 2019.
- [36] SIASS UNAM. Sistema de Información Automatizada de Servicio Social (Social Service Information System), 2019.
- [37] Jari Vanhanen, Timo OA Lehtinen, and Casper Lassenius. Software engineering problems and their relationship to perceived learning and customer satisfaction on a software capstone project. *Journal of Systems and Software*, 137:50–66, 2018.
- [38] Elaine Venson, Rejane Figueiredo, Wander Silva, and Luiz Ribeiro. Academy-industry collaboration and the effects of the involvement of undergraduate students in real world activities. In *Frontiers in Education Conference (FIE)*, 1–8, IEEE, 2016.
- [39] Ira Weissberger, Abrar Qureshi, Assad Chowhan, Ethan Collins, and Dakota Gallimore. Incorporating software maintenance in a senior capstone project. *International Journal of Cyber Society and Education*, 8(1):31–38, 2015. <http://dx.doi.org/10.7903/ijcse.1238>
- [40] Claes Wohlin and Björn Regnell. Achieving industrial relevance in software engineering education. In *12th Conference on Software Engineering Education and Training (Cat. No. PR00131)*, 16–25, IEEE, 1999.
- [41] Murat Yilmaz, Faris Serdar Tasel, Ulas Gulec, and Ugur Sopaoglu. Towards a process management life-cycle model for graduation projects in computer engineering. *PLOS ONE*, 13(11):1–17, 11 2018. <https://doi.org/10.1371/journal.pone.0208012>

## Playing the Sprint Retrospective

Maciej Wawryk

Gdansk University of Technology, Faculty of  
Electronics, Telecommunications and Informatics  
Narutowicza 11/12, 80-233 Gdansk, Poland.  
Email: wawryk2@gmail.com

Yen Ying Ng

Gdansk University of Technology, Faculty of  
Electronics, Telecommunications and Informatics  
Narutowicza 11/12, 80-233 Gdansk, Poland.  
Department of English Studies, Nicolaus Copernicus  
University, Torun, Poland.  
Email: nyysang@gmail.com

**Abstract**— In this paper, we report on a replication of the study by Przybyłek & Kotecka [2017]. The aim of our study was to revise the work practices related to Sprint Retrospectives in Bluebay Polska Sp. z.o.o. by adopting collaborative games. The feedback received from two Scrum teams confirms the findings from the original study and indicates that collaborative games improve participants' creativity, involvement, and communication as well as produce better results than the standard retrospective.

### INTRODUCTION

AGILE methods appeared as a reaction to traditional ways of developing software and acknowledged that customers are unable to definitively state their needs up front [Przybyłek, 2014; Przybyłek & Zakrzewski, 2018]. In agile software development requirements and solutions evolve through the collaboration of all stakeholders. The Agile Manifesto [Highsmith & Fowler, 2001] advocates principles and values such as face-to-face conversation within a development team, motivated individuals, self-organizing teams, and retrospectives at regular intervals. Furthermore, agile team members are expected to be proactive and creative in resolving complex software development problems [Highsmith & Cockburn, 2001; Crawford et al., 2012; Przybyłek & Zakrzewski, 2018; Przybyłek & Kowalski, 2018; Jarzębowski & Ślesiński, 2018; Miler & Gaida, 2019; Zakrzewski et al., 2019]. However, agile methods do not provide techniques to promote these attitudes. Responding to this challenge, Przybyłek and his team [Przybyłek & Olszewski, 2016; Przybyłek & Kotecka, 2017; Przybyłek & Zakrzewski, 2018; Przybyłek & Kowalski, 2018; Zakrzewski et al., 2019] proposed to equip agile teams with collaborative games.

Collaborative games refer to structured techniques inspired by game play, but designed for the purpose of solving practical problems [Przybyłek & Kotecka, 2017], for example they are quite widely used as a requirements elicitation technique [Marciniak & Jarzębowski 2016, Przybyłek & Zakrzewski, 2018]. By involving visual activities like moving sticky notes and drawing pictures, they leverage multiple dimensions of communication, which results in deeper, richer and more meaningful exchanges of

information [Przybyłek & Kowalski, 2018; Hohmann, 2006]. Besides, various studies suggest that fun is a powerful tool in unleashing creativity and facilitating collaboration [Hohmann, 2006; Trujillo et al., 2014; Ghanbari et al., 2015].

Przybyłek & Kotecka [2017] demonstrated that the promised benefits of collaborative games were materialized when running a game-based retrospective in 3 teams in Intel. The Sprint Retrospective is a meeting in which the team inspects and adapts its way of working [Ilyés, 2019]. Its purpose is to recognize the successes and failures of the last Sprint and to link the related experience to action proposals for improvements. In this paper, we report on a replication of the study conducted in Intel [Przybyłek & Kotecka, 2017]. The feedback received from 2 Scrum teams confirms the findings from the original study and indicates that collaborative games improve participants' creativity, motivation, communication, knowledge sharing, make participants more willing to attend Scrum meetings, and produce better results than the standard retrospective.

The remainder of this paper is organized as follows. Section II provides an overview of the previous studies. Section III explains the employed research methodology. Section IV reports the research project and its results. Finally, the last section concludes the paper.

### RELATED WORK

Recently, there has been lots of interest in adopting collaborative games to aid agile teams. Przybyłek & Olszewski [2016] defined an extension to Open Kanban, which consists of 12 collaborative games broken down into four categories in accordance with four Open Kanban principles. The extension was proved to assist unskilled team members better understand the principles of Kanban and promote the teamwork.

Przybyłek & Zakrzewski [2018] proposed a framework for extending Scrum with 9 collaborative games. The framework was proved to boost agile requirements engineering.

Przybyłek & Kowalski [2018] developed a web portal which provides 8 collaborative games to be used in agile software development.

Przybyłek & Kotecka [2017] adopted 5 collaborative games to support running an effective and enjoyable retrospective meetings. Our study is a continuation of their work, since we evaluate these games in other company and teams.

#### RESEARCH METHOD

Our study was conducted as Action Research [Baskerville & Myers, 2004]. In Action Research, the researcher works in close collaboration with a group of practitioners, acting as a facilitator, to solve a real-world problem while simultaneously expanding scientific knowledge [Przybyłek & Zakrzewski, 2018]. The researcher brings his knowledge of action research while the practitioners bring their practical knowledge and context [Baskerville & Myers, 2004]. A precondition for Action Research is to have a problem owner willing to collaborate to identify a problem, engage in an effort to solve it, analyze the results, and determine future actions [Przybyłek & Zakrzewski, 2018]. The problem owner in this research was Bluebay Polska Sp. z.o.o.. The company was interested in auditing its work practices related to Sprint Retrospectives and improving identified deficiencies. Two Scrum teams participated in the study (Table I). Team 1 developed a web store for Aclari Diamonds, which is a jewellery company, while Team 2 developed print management software for POSperita, which is a printer & advertising agency.

TABLE I.  
PARTICIPATING TEAMS (ROLE, EXPERIENCE IN YEARS)

Team 1	Team 2
Team Leader & Scrum Master, 10	Team Leader & Scrum Master, 10
Developer, 5	Developer, 8
Developer, 3	Developer, 6
Tester, 2	Developer, 5
	Tester, 5

#### ACTION RESEARCH IN BLUEBAY POLSKA

We identified that our teams encountered similar problems related to Sprint Retrospective as those presented in the original study [Przybyłek & Kotecka, 2017]. Accordingly, we decided to implement all the games except Mad/Sad/Glad, which was depreciated in the original study and revised by Mood++. In addition, we decided to try one new game, i.e. 360° Appreciation.

360° Appreciation [Caroli & Caetano, 2016] is a game to foster a conducive working environment that strengthens people relationship and increases team morale. It allows open positive feedback within a team as well as appreciating the time and energy spent by the team members. In other words, it focuses only on the developers' strength instead of

their weaknesses, which can be rather discouraging. The game is not complicated as it can be conducted in any environment. What is more, no additional equipments such as blackboards, posters and sticky notes are required. To run this activity, the facilitator asks everyone to write down their appreciations about one another on a piece of paper. After that, the team is asked to form a circle with one participant sitting in the middle. The other participants are asked to read their appreciation feedback to the one in the center. The same process is repeated until everybody in the team has received feedback.

Each game was implemented twice in each team. Before a game was run for the first time, it was presented to the team. After each game session, we issued a questionnaire to collect feedback from the participants. The responses were made on a five-point Likert scale. Finally, the results were analyzed and discussed with the participants.

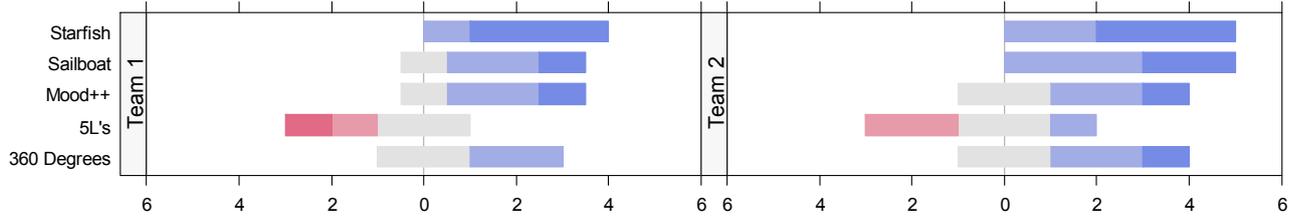
All games except 5L's and 360° appreciation were evaluated positively with respect to all categories. Playing 5L's consumed too much time, while the obtained results were worse when compared to Starfish, Sailboat or Mood++. As for 360° appreciation, even though it got low scores for questions 3-6, it is still successful overall, because it was not designed to promote these issues. The game was considered helpful in relieving the tension or getting to know new team members. Since this game does not provide any feedback on the issues during the Sprint, it should be combined with other collaborative game when performed during the retrospective. In turn, Sailboat was especially appreciated for allowing participants to identify risks in a project. The detail results are presented in Fig. 1.

#### CONCLUSIONS

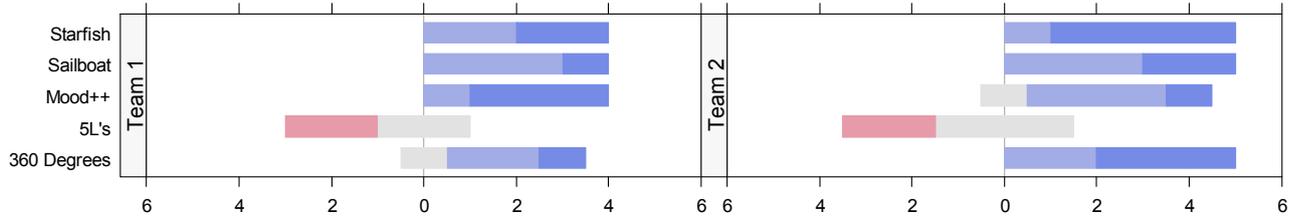
This paper reports on an Action Research project in conducted in Bluebay Polska Sp. z.o.o. Following the best practices developed by Przybyłek & Kotecka [2017], we revitalized retrospectives by adopting collaborative games. The feedback gathered from two Scrum teams confirms the previous findings that game-based retrospectives produce better results than standard retrospectives and lead to a variety of measurable societal outcomes. In particular, Starfish, Sailboat, and Mood++ improved participants' creativity, motivation, communication, knowledge sharing and make participants more willing to attend Scrum meetings.

As future work, we want to measure in a quantitative experiment with settings similar to [Przybyłek, 2018] whether game-based retrospectives are more effective than standard retrospectives. Moreover, after collecting more data, we plan to build a recommender system [Karpus, 2019] that will help scrum teams to choose a retrospective game suitable for a given context.

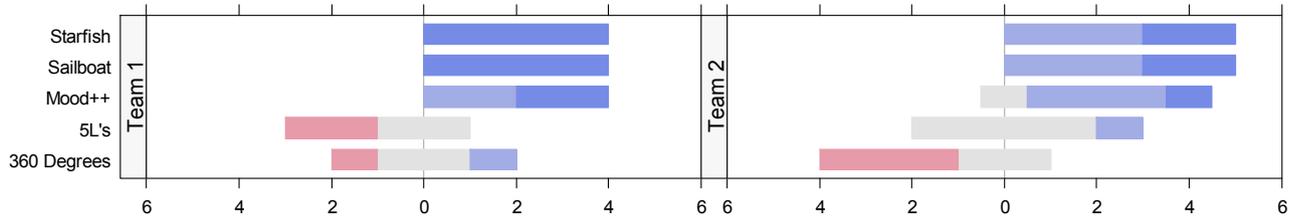
**1. The game produces better results than the standard approach**



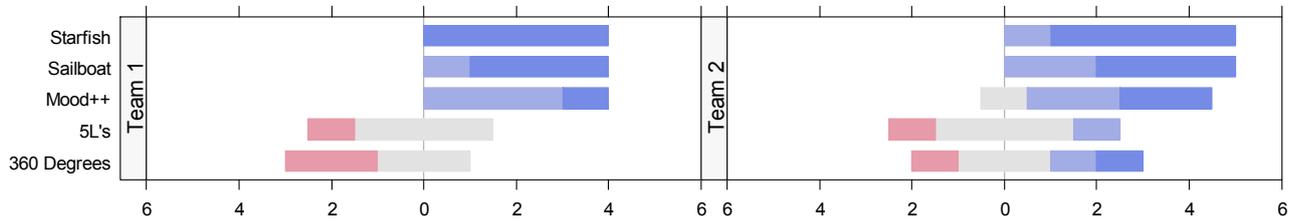
**2. The game should be permanently adopted by your team**



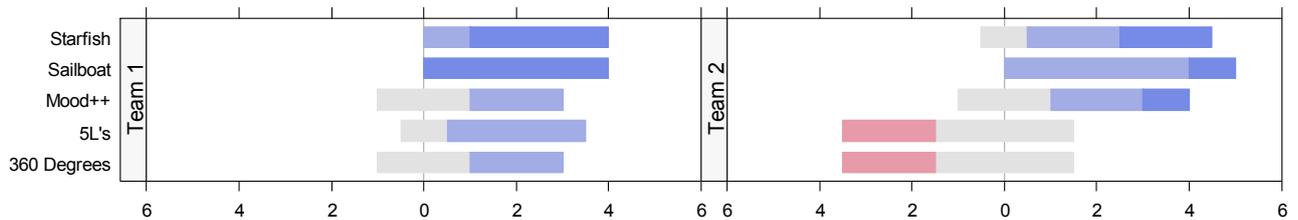
**3. The game fosters participants' creativity**



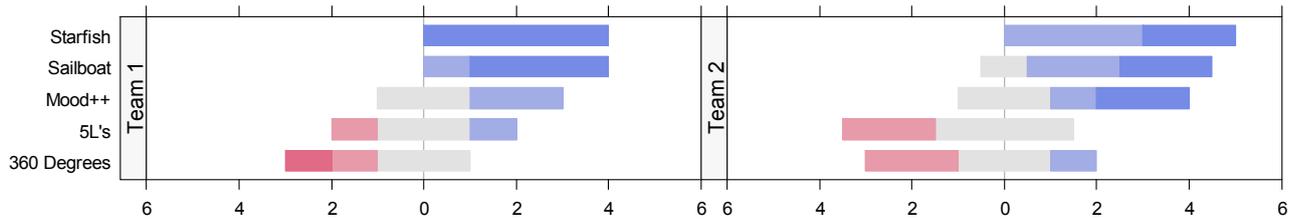
**4. The game fosters participants' motivation and involvement**



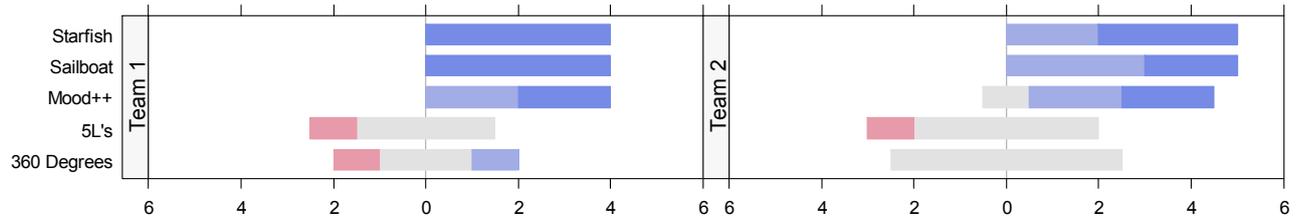
**5. The game improves communication among the team members**



**6. The game facilitates knowledge sharing among the participants**



### 7. The game makes participants more willing to attend the meeting



### 8. The game is easy to understand and play

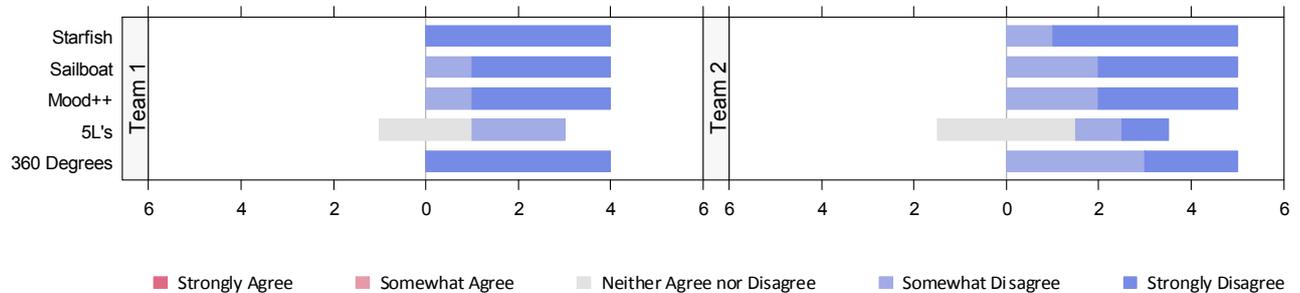


Figure 1. Aggregated results

### REFERENCES

- Baskerville, R., Myers, M.D.: Special issue on action research in information systems: making IS research relevant to practice—foreword. In: *MIS Quart* 28(3), pp. 329–335, 2004
- Caroli, P., Caetano, T.: *Fun Retrospectives - Activities and ideas for making agile retrospectives more engaging*. Leanpub, 2016
- Crawford, B., León de la Barra, C., Soto, R., Monfroy, E.: Agile software engineering as creative work. In: *5th International Workshop on Cooperative and Human Aspects of Software Engineering*, Zürich, Switzerland, 2012
- Ghanbari, H., Similä, J., Markkula, J.: Utilizing online serious games to facilitate distributed requirements elicitation. In: *Journal of Systems and Softwar*, vol. 109 (November 2015), pp. 32–49
- Highsmith, J., Cockburn, A.: Agile Software Development: The Business of Innovation. In: *IEEE Computer*, vol. 34(9), pp. 120–122, Sep., 2001
- Highsmith, J., Fowler, M.: The agile manifesto. In: *Softw. Dev. Mag.* 9, pp. 29–30, 2001
- Hohmann, L.: *Innovation Games: Creating Breakthrough Products Through Collaborative Play*. Addison-Wesley Professional, 2006
- Ilyés, E.: Create your own agile methodology for your research and development team. In: *2019 Federated Conference on Computer Science and Information Systems (FedCSIS'19)*, Leipzig, Germany, 2019
- Jarzębowicz, A., Ślesiński, W.: Assessing Effectiveness of Recommendations to Requirements-Related Problems through Interviews with Experts. In: *2018 Federated Conference on Computer Science and Information Systems (FedCSIS'18)*, Poznan, Poland, 2018
- Karpus, A., Raczyńska, M., Przybyłek, A.: Things You Might Not Know about the k-Nearest Neighbors Algorithm. In: *11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, Vienna, 2019
- Marciniak P., Jarzębowicz A.: An Industrial Survey on Business Analysis Problems and Solutions. In: *Software Engineering: Challenges and Solutions*, pp. 163-176, *Advances in Intelligent Systems and Computing* vol. 504, Springer International Publishing, ([https://doi.org/10.1007/978-3-319-43606-7\\_12](https://doi.org/10.1007/978-3-319-43606-7_12)), 2016
- Miler, J., Gaida, P.: On the Agile Mindset of an Effective Team – an Industrial Opinion Survey. In: *2019 Federated Conference on Computer Science and Information Systems (FedCSIS'19)*, Leipzig, Germany, 2019
- Przybyłek, A.: A Business-Oriented Approach to Requirements Elicitation. In: *9th International Conference on Evaluation of Novel Approaches to Software Engineering*, Lisbon, Portugal, 2014, doi: 10.5220/0004887701520163
- Przybyłek, A.: An empirical study on the impact of AspectJ on software evolvability. In: *Empirical Software Engineering*, vol. 23(4), pp. 2018–2050, August 2018, <https://doi.org/10.1007/s10664-017-9580-7>, 2018
- Przybyłek, A., Olszewski, M.: Adopting collaborative games into Open Kanban. In: *2016 Federated Conference on Computer Science and Information Systems (FedCSIS'16)*, Gdansk, Poland, 2016, doi: 10.15439/2016F509
- Przybyłek, A., Kotecka, D.: Making agile retrospectives more awesome. In: *2017 Federated Conference on Computer Science and Information Systems (FedCSIS'17)*, Prague, Czech Republic, 2017, doi: 10.15439/2017F423
- Przybyłek, A., Kowalski, W.: Utilizing online collaborative games to facilitate Agile Software Development. In: *2018 Federated Conference on Computer Science and Information Systems (FedCSIS'18)*, Poznan, Poland, 2018, doi: 10.15439/2018F347
- Przybyłek, A., Zakrzewski, M.: Adopting Collaborative Games into Agile Requirements Engineering. In: *13th International Conference on Evaluation of Novel Approaches to Software Engineering (ENASE'18)*, Funchal, Madeira, Portugal, 2018, doi: 10.5220/0006681900540064
- Trujillo, M.M., Oktaba, H., González, J.C.: Improving Software Projects Inception Phase Using Games: ActiveAction Workshop. In: *9th International Conference on Evaluation of Novel Approaches to Software Engineering (ENASE'14)*, Lisbon, Portugal, 2014
- Zakrzewski, M., Kotecka, D., Ng, Y.Y., Przybyłek, A.: Adopting Collaborative Games into Agile Software Development. In: *Damiani E., Spanoudakis G., Maciaszek L. (eds) Evaluation of Novel Approaches to Software Engineering. ENASE 2018. Communications in Computer and Information Science*, vol 1023. Springer, Cham, 2019, doi: 10.1007/978-3-030-22559-9\_6

# Security-oriented agile approach with AgileSafe and OWASP ASVS

Katarzyna Łukasiewicz  
Gdańsk University of Technology  
ul. Narutowicza 11/12, 80-233  
Gdańsk, Poland  
Email: katlukas@pg.edu.pl

Sara Cygańska  
IHS Markit  
ul. Marynarki Polskiej 163, 80-  
868 Gdańsk, Poland  
Email:  
sara.cyganska@ihsmarkit.com

**Abstract**—In this paper we demonstrate a security enhancing approach based on a method called AgileSafe that can be adapted to support the introduction of OWASP ASVS compliant practices focused on improving security level to the agile software development process. We also present results of the survey evaluating selected agile inspired security practices that can be incorporated into an agile process. Based on the survey's results, these practices were used as an input to AgileSafe method as well as to demonstrate their potential to comply with OWASP ASVS requirements.

## I. INTRODUCTION

THE concern for providing secure systems has become increasingly important throughout the years. With the rapid progress in the IT domain, expansion of the internet solutions and the level of general computer science knowledge, the problem with security affects multiple domains. At the same time, the changing markets and need for flexibility encourages many companies to adopt agile approach [1].

The goal of the research described in this paper was to identify security-focused agile practices, evaluate their usability and impact so that the positively assessed practices could be incorporated into an OWASP ASVS [2] compliant process, as a part of AgileSafe method [3].

## II. BACKGROUND

### A. Agile methods

Ever since the announcement of the Agile Manifesto [4], the agile methods such as Scrum [5], eXtreme Programming [6] or Kanban [7] have been growing increasingly in popularity. The reports of the benefits experienced by numerous companies [8][9] encouraged the trend to shift from traditional, plan-driven methods to the agile ones. What is important is that this shift has not only concerned small and evolving companies which are considered a target of the agile approach. Bigger organizations with larger teams or corporate structures have also sought ways to incorporate agile approach, which resulted in methods such as SAFe [10] or DevOps [11].

### B. OWASP ASVS

The name of the OWASP Application Security Verification Standard (OWASP ASVS) comes from the organization with same name, which created it - The Open Web Application Security Project [12]. Its two main goals are to help creating and maintaining secure software and help in defining requirements between service providers and their clients.

OWASP ASVS has been chosen for this research due to its versatility, open access and popularity among practitioners [13]. The domain of web applications is at the forefront of security issues, with frequent news about major security breaches [14]. For this reason, catering a solution that would allow combining agile security practices with OWASP ASVS requirements could be of interest to many organizations.

## III. AGILESAFE

In the safety context, quite similarly to the security one, norms and standards are vital to ensure the level of trust and quality of high-integrity systems. In order to enable safety-critical software companies to adopt hybrid agile approach while satisfying the regulatory requirements of applicable standards, AgileSafe [15] method has been proposed. It presents a framework for collecting and suggesting the most suitable agile practices for a given project, as well as the means for managing and monitoring conformance with the applicable regulatory requirements.

### A. Overview

As an input to AgileSafe takes the characteristics of a project in which the new approach will be implemented (Project Characteristics) as well as a list of regulations (Regulatory Requirements), which the project needs to comply with.

Based on this information, the user is guided through the process of practices suggestion as well as the process of preparing a set of assurance arguments [16] that will help the user to maintain conformance with given norms and standards. As a result, the user obtains a tailored Project Practices Set, which would best suit a project with given characteristics and regulation restrictions as well as a set of

assurance arguments to monitor compliance with the chosen regulations.

### B. Practices Knowledge Base

The information about practices available in AgileSafe, their capability to answer given Project Characteristics and Regulatory Requirements, is kept in the Practices Knowledge Base. Each practice is described using the same template that is then translated into OWL and managed using Protégé [17].

### A. Assurance arguments

In order to ensure that the Regulatory Requirements will be met when applying the new agile approach, AgileSafe uses a set of assurance arguments. The highest level of abstraction is represented by Practices Compliance Argument. It is created separately for each standard added to the method and collects all of the practices from Practices Knowledge Base that have a potential to answer the standard's requirements. Such practices are arranged accordingly in the argument structure for a given standard requirements.

In this particular research, we focused on the most general Practices Compliance Argument for OWASP ASVS and the security-oriented practices identified in the course of this research, to keep it independent from any particular software project.

## II. SECURITY-ORIENTED AGILE PRACTICES

In order to propose agile security practices that could extend the Practices Knowledge Base of the AgileSafe method, a review of the scientific literature and articles on blogs and industry portals was carried out.

### A. Identification of security-oriented agile practices

While there are many well-known security-oriented practices such as threat modelling or attack trees, in this research we wanted to expand this list and focus on less obvious, agile inspired practices, to enrich the Practice Knowledge Base of AgileSafe method.

A literature review has been performed and as a result 10 articles were selected to be used in further work [18][19][20][21][22][23][24][25][26][27].

### B. Selected practices description

Based on the articles identified in the research, 10 hybrid agile security-oriented practices were identified:

*Abuser Stories.* They describe, using a form similar to regular User Stories, how the system might be attacked and how assets might be put in risk. They should be estimated in accordance to how much damage they may potentially cause and probability of a successful attack. [19]

*Evil user stories.* This practice describes actions of malicious user (e.g. "As a hacker I want to steal payment information of other clients, so I can sell it."). They may be used as a starting point for threat modelling. [20]

*Misuse cases.* They are negative use cases. They illustrate behavior not wanted in the system, that can cause a security breach and can be described using UML diagrams. [21]

*Protection poker.* This is a software security game intended to create a list of each requirement relative security risk. It derives from Planning Poker technique of estimation. [22]

*Second delivery.* This is a process, that aims to integrate security related solutions to the project that already satisfies functional requirements. It is based on XP methodology. [23]

*Security engineer.* It calls for adding an expert role, that brings up-to-date security knowledge to developers' team. His insight is useful during multiple phases and actions in project.

*Security Sprint.* This is a practice inspired by Scrum. It's similar to regular Sprint except that it focuses on security issues. [24]

*Security-focused code reviews.* Such reviews should be performed for every story separately – no story can be completed without security review, fixing findings from review and then passing re-review. [25]

*S-Mark and S-Tag.* Originating from Secure Scrum, they are a way to document identified security issues in Scrum Backlog by creating system of tags (security issues) and markings for stories related to respective tags. [18][27]

*Spikes.* They are a way to include security analysis and design within Scrum. They accommodate activities that don't produce customer-valued product, like security analysis or system designing. [26]

## III. SURVEY

In order to evaluate the usability and accessibility of the selected security-oriented agile practices in projects with high security requirements a survey was conducted. It tackled 10 specific agile security-oriented practices, asking the respondents to rate their respective ease of use and security enhancement potential.

Subjects chosen to participate in the survey were 24 IT practitioners (both development and operations) from 7 different software companies, ranging from small to corporate ones, from Poland and UK. The questionnaire was distributed mostly by email and direct messages in social networks, eliminating probability of acquiring responses from random, unrelated to the field respondents. The respondents were also provided with the practices detailed descriptions.

### A. Results

For each practice two closed questions were asked about its ease of use and if it's improving security in the project. In total, 15 of all the participants made their choices in those questions. Also, each practice was open to comments from the respondents. The results are presented in the Fig. 3 and Fig. 4.

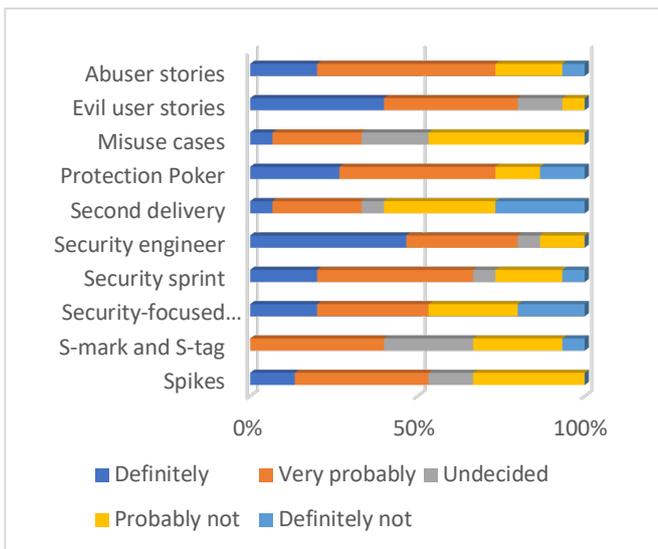


Fig. 3 Is this practice ease to use?

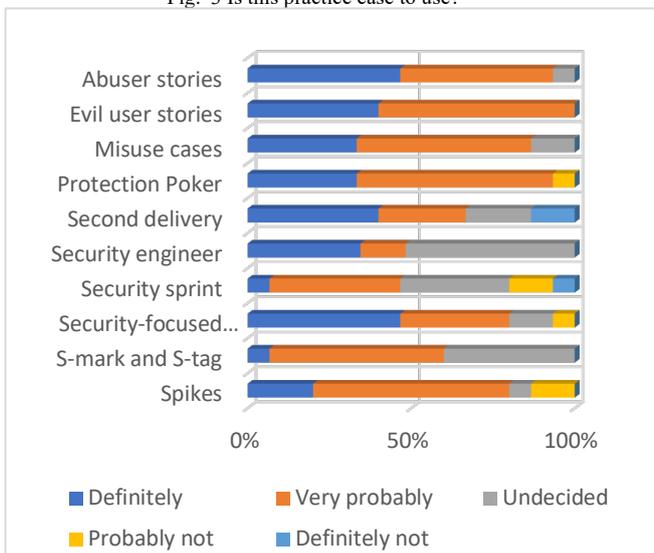


Fig. 1 Does this practice improve security in the project?

**Abuser Stories.** None of respondents chose negative answer for this practice security improvement potential and not many had doubts about its positive influence. But 26,67% believed it would not be easy to use - as the reason they mostly described difficulty in estimating attack probability. Despite this fact, this practice has potential benefits in the projects wanting to comply with OWASP.

**Evil user stories.** This practice was also positively rated in terms of security improvement. What's more, only 20% expressed doubts or were undecided about its ease of use. Those results categorize it as both efficient and easy to get started with. Respondent commented on possible threat to project agility in case of creating a large number of evil user stories.

**Protection Poker.** Majority of respondents found this practice easy to use – among the benefits they listed possible automation of prioritization. The doubts were similar to those for Abuser Stories practice – difficulty in

estimation of attack ease and probability. Another noticed difficulty is the necessity for security experts to participate in the process. Despite that problems only 7% didn't rate the practice positively in terms of security.

**Second delivery.** This practice didn't occur as easy to use to most respondents. A lot of them were concerned about the need to re-implement huge parts of system in order to satisfy security requirements. 67% of answers in question about security were positive, but considering its difficulty, this practice might not cause some problems in actual development process. Also, a significant problem with security was noticed, that during the first development unexpected security flaws might be introduced to the system that are not addressed in the second delivery.

**Security engineer.** Most of respondents rated this practice positively in terms of ease of use, as it wouldn't require additional amount of work from the team and it would be beneficent to have an expert that is not writing the code himself. Among listed problems were difficulty in finding the suitable person for this role and risk of putting all of responsibility for security on one person. Despite those issues, rating in security improvement area was positive, with only 7% of participant undecided and none rating it negatively.

**Security Sprint.** The majority of respondents rated this practice as easy to use, but doubts were expressed that it could lead to development work duplications. Also, the question was asked about the case in which not enough security tasks are defined to fill the whole sprint. 47% of answers were positive in terms of security improvements, but as much as 33% of participants were undecided. This can indicate that practice description should be clarified when added to the AgileSafe Knowledge Base.

**Security-focused code reviews.** Opinions on this practice's ease of use are divided – the results for "Definitely" and "Definitely not" are equal (20%). Among mentioned problems were difficulty with finding a suitable expert and a lot of additional effort required for conducting such reviews. Despite that, most of respondents decided that this practice improves security in the project (80%). But the expected improvement seems not to be worth the effort required.

**S-Marks and S-Tags.** None of the respondents found this practice definitely easy to use, and 40% decided it's probably easy to use. Considering amount of answers "Undecided" in both questions, this practice might be too complicated to take up without previous training. Practice gained no negative rating in terms of security, but concerns were raised that it might be possible to lose track of some tags and marks and therefore omit some security issues in development. Also, the question was asked about support in existing project management tools, which could solve tracking problem.

**Spikes.** Although the majority of respondents (53%) rated this practice as easy to use, 33% doubted it – some commented that it's difficult to understand. However, in terms of security, most of participants expressed no concern

about its influence on project security. A question was also asked about other practices that can be used in security projects development. Only two answers were provided – bug bounty and security hackathon. This shows that it's not a common knowledge among developers.

The results show that, although not all practices are easy to use, most of them serve their purpose well by explicitly requiring some security assurance activities. Some of those that scored lowest in terms of easiness might be improved by description clarification, training or providing supporting tools.

#### IV. OWASP ASSURANCE ARGUMENT

Because of the positive results of practices security assurance evaluation, the next step was to add them to the Practices Knowledge Base. The selected practices were analyzed according to the AgileSafe practice description template and incorporated into the knowledge base. Newly added security practices were assessed with respect to their OWASP conformance potential.

OWASP ASVS requirements has been added to the method and based on the Practices Compliance Assurance Argument Pattern, were mapped to the Practices Compliance Assurance Argument using NOR-STA tool [28].

All of the OWASP ASVS requirements were successfully mapped into the structure. The practices that were able to answer specific requirements were attached with a relevant rationale in the NOR-STA tool. None of the requirements were left without a practice that might be able to provide conformance.

It is worth noting that there was not one practice that would sufficiently address all of the OWASP ASVS requirements, which means that in a project wishing to comply with the standard, implementing a combination of the analyzed practices would be needed.

The prepared Practices Compliance Argument has been accepted as a part of the AgileSafe potential extension for security assurance domain. Based on this argument, depending on a given project's Project Characteristics, a new hybrid approach with OWASP ASVS compliance potential could be suggested.

#### V. CONCLUSIONS

During the literature review, 10 security-oriented agile practices were identified. The practices were positively assessed in the conducted surveys and successfully enriched the Agile Practices Knowledge Base. The OWASP ASVS was mapped into the method and formed, along with the identified practices, the Practices Compliance Argument, which after updating it with all of the other applicable practices available in AgileSafe, might be further used to support practices selection in specific projects. A case study carried out with such projects, going through the whole practices selection process of AgileSafe might be performed as next step of the research.

#### REFERENCES

- [1] "VersionOne® Releases 11th Annual State of Agile Report", VersionOne, 2017. [Online]. Available: <https://www.versionone.com/about/press-releases/versionone-releases-11th-annual-state-of-agile-report/>
- [2] J. Manico, "OWASP Application Security Verification Standard," 2015.
- [3] K. Łukasiewicz, J. Górski, "AgileSafe – a method of introducing agile practices into safety-critical software development processes," *Proceedings of the Federated Conference on Computer Science*, vol. Vol. 8, pp. 1549-1552, 2016.
- [4] Agile Manifesto., *Manifesto for Agile Software Development*. 2001 [online] Available at: <http://agilemanifesto.org>.
- [5] K. Schwaber and M. Beedle, *Agile software development with scrum*. Upper Saddle River, N.J: Prentice Hall, 2002
- [6] K. Beck and C. Andres, *Extreme programming explained*. Addison-Wesley Professional, 2004.
- [7] D. Anderson, *Kanban*. Sequim: Blue Hole Press, 2010.
- [8] J. Drobka, D. Noftz and R. Raghu, "Piloting XP on four mission-critical projects". *IEEE Softw.*, 21(6), pp.70-75, 2004
- [9] M. Lindvall., D. Muthig, A/ Dagnino, C. Wallin, M. Stupperich, D. Kiefer, J. May & T. Kähkönen. "Agile Software Development in Large Organizations" in *Computer*, 37(12), pp. 26-34, 2004.
- [10] R. Knaster, D. Leffingwell, *SAFe Distilled: Applying the Scaled Agile Framework for Lean Software and Systems Engineering*. Addison-Wesley Professional, 2017.
- [11] J Kim, G., Willis, J., Debois, P., Humble, J., Allspaw, J. *The DevOps Handbook*. Trade Select, 2016.
- [12] OWASP, "OWASP," [Online]. Available [https://www.owasp.org/index.php/Main\\_Page](https://www.owasp.org/index.php/Main_Page).
- [13] OWASP users [Online] Available: [https://www.owasp.org/index.php/Category:OWASP\\_Application\\_Security\\_Verification\\_Standard\\_Project#tab=ASVS\\_Users](https://www.owasp.org/index.php/Category:OWASP_Application_Security_Verification_Standard_Project#tab=ASVS_Users)
- [14] World's Biggest Data Breaches & Hacks, 2019, [Online] Available: <https://www.informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks/>.
- [15] K. Łukasiewicz "Method of selecting programming practices for the safety-critical software development projects," Ph.D. dissertation, Dept. Soft. Eng., Gdańsk Univ. of Technology, Gdańsk, Poland, 2019.
- [16] J Górski, J., Jarzębowski, A., Leszczyna, R., Miler, J. and Olszewski, M. "Trust case: justifying trust in an IT solution". *Reliability Engineering & System Safety*, 89(1), pp.33-47. 2005
- [17] Musen, M.A. "The Protégé project: A look back and a look forward". *AI Matters*. Association of Computing Machinery Specific Interest Group in Artificial Intelligence, 1(4), June 2015.
- [18] J D. Mougouei, N. Fazlida, M. Sani, M. M. Almasi, "S-Scrum: a Secure Methodology for Agile Development of Web Services," *World of Computer Science and Information Technology Journal (WCSIT)*, vol. 3, no. 1, pp. 15-19, 2013.
- [19] J. Peeters, "Agile security requirements engineering." *Symposium on Requirements Engineering for Information Security*, 2005
- [20] E. A. Fischer, "Federal Laws Relating to Cybersecurity: Overview of Major Issues, Current Laws, and Proposed Legislation," 2014
- [21] G. Sindre, A. L. Opdahl, "Eliciting security requirements with misuse cases".
- [22] L. Williams, A. Meneely, G. Shipley, "Protection Poker: The New Software Security "Game"".
- [23] E. G. Aydal, R. F. Paige, H. Chivers, P. J. Brooke, "Security Planning and Refactoring in Extreme Programming"
- [24] G. Boström, J. Wäyrynen, M. Bodén, K. Beznosov, P. Kruchten, "Extending XP Practices to Support Security Requirements Engineering"
- [25] T. Nguyen, "Integrating Security into Agile Methodologies," <http://www.umsl.edu/~sauterv/analysis/F2015/Integrating%20Security%20into%20Agile%20methodologies.html>
- [26] OWASP, "Agile Software Development: Don't Forget EVIL User Stories," [https://www.owasp.org/index.php/Agile\\_Software\\_Development:\\_Don%27t\\_Forget\\_EVIL\\_User\\_Stories](https://www.owasp.org/index.php/Agile_Software_Development:_Don%27t_Forget_EVIL_User_Stories).
- [27] C. Pohl, H.-J. Hof, "Secure Scrum: Development of Secure Software with Scrum," in *SECURWARE 2015 : The Ninth International Conference on Emerging Security Information, Systems and Technologies*, 2015
- [28] NOR-STA project Portal . 2017. [online] Available at: [www.nor-sta.eu](http://www.nor-sta.eu)

# 7<sup>th</sup> Conference on Multimedia, Interaction, Design and Innovation

**M**IDI Conference provides an interdisciplinary forum for academics, designers and practitioners to discuss the challenges and opportunities for enriching human interaction with digital products and services.

The main focus of MIDI Conference is exploring design methods for creating novel human-system interaction, developing user interfaces and implementing innovations in user-centred development of advanced IT systems and on-line services.

## TOPICS

Topics of interest include (but are not limited to) the following areas:

- interactive multimedia and multimodal interaction design
- novel interaction techniques, voice interfaces, interactive multimedia
- ubiquitous, multimodal, pervasive and mobile interaction, wearable computing
- novel information visualization and presentation techniques, Augmented/Virtual Reality
- design methods for usability, accessibility and outstanding user experience
- prototyping of user interfaces and interactive services
- human-centred design practices, methods and tools, user interface design
- unfolding trends in HCI research and practice, customer experience, Service Design
- advances in user-centred interaction design
- understanding people and interactions: theory, concepts, models and methods
- understanding people and interactions: contextual, ethnographical and field studies
- critique and evolution of methods, processes, theories and tools for human-computer interaction
- novel methodologies for conceptualization, design and evaluation of interactive products and services

## EVENT CHAIRS

- **Marasek, Krzysztof**, Polish-Japanese Academy of Information Technology, Warsaw, Poland, Poland
- **Romanowski, Andrzej**, Lodz University of Technology, Poland, Poland

- **Sikorski, Marcin**, Gdansk University of Technology, Gdansk, Poland, Poland

## PROGRAM COMMITTEE

- **Biele, Cezary**, Information Processing Institute, Warsaw, Poland, Poland
- **Forbrig, Peter**, University of Rostock, Germany, Germany
- **Guttormsen, Sissel**, University of Bern, Institute of Medical Education, Switzerland, Switzerland
- **Korżinek, Danijel**, Polish-Japanese Academy of Information Technology, Warsaw, Poland, Poland
- **Kołąkowska, Agata**, Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, Poland, Poland
- **Landowska, Agnieszka**, Gdansk University of Technology, Poland, Poland
- **Masoodian, Masood**, Aalto University, Finland, Finland
- **Miler, Jakub**, Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, Poland, Poland
- **Obaid, Mohammad**, Uppsala University, Sweden, Sweden
- **Satalecka, Ewa**, Polish-Japanese Academy of Information Technology, Warsaw, Poland, Poland
- **Slavik, Pavel**, Czech Technical University, Prague, Czech Republic, Czech Republic
- **Szklanny, Krzysztof**, Polish-Japanese Academy of Information Technology, Warsaw, Poland, Poland
- **Wichrowski, Marcin**, Polish-Japanese Academy of Information Technology, Warsaw, Poland, Poland
- **Wieczorkowska, Alicja**, Polish-Japanese Academy of Information Technology, Warsaw, Poland, Poland
- **Winkler, Marco**, Université Nice Sophia Antipolis, France, France
- **Wojciechowski, Adam**, Institute of Information Technology, Lodz Univ. of Technology, Poland, Poland
- **Wołk, Krzysztof**, Polish-Japanese Academy of Information Technology, Warsaw, Poland, Poland
- **Ziegler, Juergen**, University of Duisburg-Essen, Germany, Germany



# Exploration of older drivers interaction with conversation assistant

Jakub Berka, Lukas Chvatal, Zdenek Mikovec

Faculty of Electrical Engineering,  
Czech Technical University in Prague  
Prague, Czech Republic

berkajak@fel.cvut.cz, chvatlu2@fel.cvut.cz, xmikovec@fel.cvut.cz

**Abstract**—The number of older drivers will be increasing therefore, their needs and requirements need to be taken into account when designing in-car user interfaces. Current solutions in car industry tend to use big touch screens for controlling the secondary tasks, such as navigation. These solutions are proved to be very distracting while driving. We present a design of conversational assistant for older drivers to improve secondary-task performance, help with decision making in primary tasks with reduced stress. We conducted a user study in the laboratory (N = 7) and gained initial knowledge about how the conversational assistant should support older drivers in secondary tasks. Our exploration revealed potential opportunities for the future design of such in-car assistants.

**Index Terms**—Driving, Voice Interfaces, Multitasking, User Centered Design

## I. INTRODUCTION

THE number of older drivers will be increasing, as the percentage population aged 65 and over was 13,8% in 2000 and 18,8% in 2017, it is expected that it will rise to 24% by the year 2030 [1]. Whatmore, seniors use individual car transport much more compared to the past [2].

With growing age drivers start to experience more frequent problems with their visual perception, attention and fast decision making [3].

Unrelated to the driver's age, all drivers are exposed to a steadily increasing number of distractors. According to [4] and [5] distractors can cause secondary task distraction, which diverts the driver's attention away from the primary task - driving. The driver becomes occupied with events which are unrelated to driving, occur away from the forward roadway and urge the driver also to look away from the forward roadway. We can divide these distractors into two groups - internal distractors (inside the vehicle) and external distractors (objects located outside the vehicle). In this paper, we are investigating only the internal distractors, which in our case are navigation system and messaging while driving. The distraction has a significant contribution to driving accidents [4] (23% of all crashes and near-crashes are caused by secondary task distraction). According to [6] the potential for a secondary task to distract the driver is determined by the task complexity, current driving demands, driver experience and skills, as well as driver's willingness to engage in the task

The influence of distracting tasks on driving performance is bigger for more complex activities, especially when drivers are

older [7]. Therefore our goal is to reduce the secondary task complexity and workload with the help of a conversational assistant. It is now possible to implement conversational assistants into cars, thanks to the progress in speech recognition systems in recent years. Also with the arrival of car-to-car and car-to-infrastructure communication, these intelligent assistants can become safer and less distracting, as they will take into account broader context, for instance, road situation, nearby traffic, etc.

In this paper, we present results from a Wizard of Oz user study with older drivers where we explored the experiences with conversational assistant designed for older drives to help them handle the secondary tasks (navigation task and messaging task) with maintaining the safety and lowest distraction as possible. Our goal was to explore how such system as a conversational assistant will be accepted by older drivers and what benefits it could bring them. Our design evoked various reactions: some drivers found the conversation with the system while driving still too distracting, but some were satisfied with our design mainly in messaging task where our proposed semi-autonomous messaging system was rated very positively.

## II. RELATED WORK

In this section, we examine existing solutions and concepts related to secondary tasks, distractions and use of conversation interaction while driving. We focused mainly on studies where the target groups were mostly older people or where the studies aimed at problems related to our case of study.

### A. Distractions while driving

Distraction can be often caused by other persons - passengers or someone on the phone, but the effect is not usually the same in both cases. Passenger is a direct participant in traffic so the conversation can be modified according to the situation on the road. In opposite, a phone call cannot be naturally suppressed according to conversation suppression hypothesis [8]. Bruyas et al. rescheduled these real-time tasks to become asynchronous. Their results show that it can help to reduce pressure on driver compared to synchronous phone communication, as a suitable place can be chosen in respect of traffic situation. Another most common problem is manual (frequently finger touches on the screen) control of infotainment systems which leads to long or frequent off-road glances

causing great danger. The study by Lee et al. [9] examines how errors in interacting with infotainment systems influence driving performance, specifically input of the words using a touchscreen, and how drivers recover from these errors. They suggest that for preventing high distraction caused by secondary task, sufficient but not greater than necessary visual information should be provided.

Although the speech-based interface has also some negative impact on driver's workload, according to Maciej and Vollrath [10] it is still better than displays and manual controls.

### B. Multimodal Interaction in the Car

One possible approach to overcome the limitations of speech-only interfaces can be multimodal interaction, which may provide fine-grained control with immediate feedback and easy undo of actions. Pfleging et al. [11] designed an interaction that combines speech and gestures on the steering wheel. To adjust distraction when interacting with infotainment systems some novel interactions techniques are becoming popular, e.g. mid-air gestures. However, the problem is with the feedback, which is still mainly through visual displays, Shakeri et al. [12] investigated different types of feedback modalities. Their study concludes that non-visual feedback (auditory, tactile) can significantly reduce distraction. In the work of Tashev et al. [13] authors created multimodal dialog system for infotainment and also formulated key requirements for voice enable infotainment systems, concerning the efficiency of multimodal interaction during high cognitive load situations.

### C. Older drivers

It is obvious that older drivers have special demands. Besides some physical impairments (visual, motion, etc.), they are more sensitive to time pressure and complexity of the tasks. On the other hand, older drivers are calmer, less reckless and less daring drivers than earlier in life and there are many aspects from which older drivers can profit. For example from their life-long driving experience, maturity and flexibility to drive at times and places that they perceive as being safer [6]. According to Bjelkemyr et al. [14] this flexibility closely relates to a phenomenon of self-regulation, which can manifest as avoiding certain conditions (e.g. driving at night or during rush hour) or difficult traffic situations (e.g. driving through specific intersections), next reducing speed, avoiding motorways, big cities, long distance travels and avoiding unknown cities, etc.

### D. Secondary tasks

There is a wide variety of secondary tasks that can be performed while driving. For our research, we have chosen two tasks: navigation on the unknown route and messaging as they are highly distracting the driver [15]. The navigation system is common equipment in cars with a lot of useful functions nowadays, but design low-distracting user interface is not an easy task, many drivers also use their mobile phones as a navigation aid. Use of mobile phones (hand-held) in

many European countries is prohibited, but drivers keep using it. Hands-free use is also proven to be distracting while driving [16]. On the other hand, there are cases in which phone calling or sending messages is necessary for drivers, therefore they need to be supported by intelligent systems to finish these tasks safely.

1) *Navigation*: The navigation task is typically supported by navigation systems that heavily rely on the visual (displays with maps) and physical (touch screen with user controls) interaction. The distraction level, especially in some stressful situation is very high [6], [10], [14]. For the purposes of an experiment, we focused on dealing with stressful error situations. According to research by Bjelkemyr et al. [14] older people reported to be calmer, less reckless and less daring drivers than earlier in life, but less busy roads are more important to them than the duration of the trip, because of their time-flexibility [6]. Bjelkemyr et al. also found out they have problems in finding their final destination and need to plan their travel in advance. Whatmore, mental states during driving can influence mood after arrival (stress), the passenger is often used as a co-pilot when in some stressful traffic situation. The researchers conclude that support systems for older drivers should increase comfort and decrease their level of stress.

2) *Messaging*: For the messaging task, there are several insights and notes that we took into account. The use of a mobile phone and even the hands-free phone calls while driving are proven by research to be very distracting and dangerous secondary tasks [16]. Atchley et al. [17] examined interview with 348 young responders, and the results show that sending, replying and reading of text messages have been recognized as riskier behavior compared to talking on the mobile phone. On the other hand, when having a mobile phone conversation driver's reaction time is increasing with the increased time of the conversation [18]. Moreover, according to Dula et al. [19] more emotional and intense phone conversations tends to cause more dangerous driving behaviors. Fofanova and Vollrath [6] also state that older people have worse driving performance when using a mobile phone. According to Lipovac et al. [16] with the growing age, the percentage of those who considered mobile phone use while driving an unsafe activity increased.

## III. DESIGN

In this section, we propose a concept of conversational assistant designed for older drivers. For the beginning of our research, we focused only at two selected tasks and the design process started with the following scenarios.

### A. Dealing with error stressful situation.

This scenario covers the recovery from error and stressful situations (see sections III-A1 and III-A2). The error situations can be identified either by the system, or they can be identified by the user itself. When the system identifies this situation, the driver will be informed about it. Then the driver can confirm or refuse that the situation is erroneous. The system then will suggest the driver how to solve the situation. Furthermore, the driver can modify the suggested solution or accept it.

TABLE I

EXAMPLE 1 OF THE DIALOG BETWEEN THE USER (U) AND THE CONVERSATIONAL ASSISTANT (A) FOR SCENARIO 1.

- A:** It seems that is hard to turn left now, is this true?  
**U:** Yes, I can't make it.  
**A:** That is no problem, you can continue straight this lane, there is another way and the delay will be only one minute.  
**U:** Okay, that's good.  
**A:** After the crossroad, try to get to the left lane when it is possible. You will be informed about next steps, don't worry.

TABLE II

EXAMPLE 2 OF THE DIALOG BETWEEN THE USER (U) AND THE CONVERSATIONAL ASSISTANT (A) FOR SCENARIO 1.

- A:** It seems that it is hard to turn left now, is this true?  
**U:** Yes, I can't make it.  
**A:** That is no problem, you can continue straight this lane, there is another way and the delay will be only one minute.  
**U:** I rather turn right, the situation in front of me looks complicated.  
**A:** Okay, turn right. There is a better way, which is more peaceful and the delay is insignificant - only 4 minutes more. You have plenty of time to get to your final destination.

1) *Wrong lane on a crossroad:* The driver is arriving at the crossroad with three lanes. Traffic is heavy, s/he is in the most right lane but according to the navigation, s/he should be in the most left lane to turn left. System evaluated that s/he is too close to the crossroad to change the lane safely (see dialog examples in I, II and III).

TABLE III

EXAMPLE 3 OF THE DIALOG BETWEEN THE USER (U) AND THE CONVERSATIONAL ASSISTANT (A) FOR SCENARIO 1.

- A:** It seems that is hard to turn left now, is this true?  
**U:** No, I just want to turn right to the fast food on the corner.  
**A:** Okay, do you mean "McDonald's" restaurant?  
**U:** Yes.  
**A:** Perfect, go ahead and the parking lot is in 200 meters. The entrance to the drive thru is few meters further.

2) *Lack of fuel:* Driver noticed that a gasoline indicator shows that there is no much gasoline in the tank, feels insecure, thus starts conversation with the system (see dialog example in IV).

### B. Semi-autonomous messaging system

For our first design iteration of messaging task we selected model situation described in the following scenario. The scenario represents situation when the driver wants to send informative sms messages about the journey to someone (e.g. a partner, relatives). See dialog example in V.

*It is Thursday morning and in a town's swimming pool ends regular lecture for seniors. One of the participants is*

TABLE IV

EXAMPLE 4 - EXCERPT FROM THE DIALOG BETWEEN THE USER (U) AND THE CONVERSATIONAL ASSISTANT (A) FOR SCENARIO 1.

- U:** I have a low amount of fuel, is it enough to get home?  
**A:** It should be enough to get home, but I recommend to stop at petrol station due to the traffic situation. Do you want to find petrol stations?  
**U:** Yes, find the stations.  
**A:** There is a petrol station OMW twelve kilometers ahead on the road you are going. Do you want to stop there?  
**U:** Nah, is there any Shell station? I have their VIP card.

TABLE V

EXCERPT FROM THE DIALOG BETWEEN THE USER (U) AND THE CONVERSATIONAL ASSISTANT (A) FOR SCENARIO 2.

- U:** I want to write an sms.  
**A:** I see that you are driving home, do you want to set sending sms to Peter (husband) or someone else?  
**U:** To Peter, please.  
**A:** Got it. Should I send him that you are about to go home?  
**U:** Yes.  
**A:** Do you want to add something to the message?  
**U:** No.  
**A:** Okay. I have sent the message to Peter. *Maria had started driving but after few kilometers she got stuck in a traffic jam.*  
**A:** It seems that there will be a little delay due to this traffic jam. Should I send a message to Peter about it?  
**U:** Yes, please.  
**A:** I will send him this text: "Hi Peter, I am in a traffic jam on the Neibolt street, it seems that I will arrive ten minutes later, around 6:45. Maria." Should I add something?'

*70-year-old Maria, who is an enthusiastic swimmer and she never misses a lecture. She lives with her husband Peter in the nearby city, but the connection by public transport is not good, so she uses car for transportation. Her husband is worried about her drive back as she could be tired after an hour of swimming. Unfortunately, there is no other option. Thankfully, they have got a semi-autonomous messaging system in their car so he can be calmer when gets messages about the journey of his wife. After the engine had started, Maria got information that the messaging system is ready and asked if she wants to send a message to her husband that she is about to go. The system reads a prepared message and asks for confirmation. During the way home, Maria got into the rush hour of a city so got stuck in a traffic jam. In that moment the system asked if she wants to inform her husband about the traffic situation and delay. After her reply, the system sent a message to Maria's husband about the situation and estimated delay and then informed her that the message had been sent. Both Maria and her husband can be relaxed, because they know that everything is alright.*

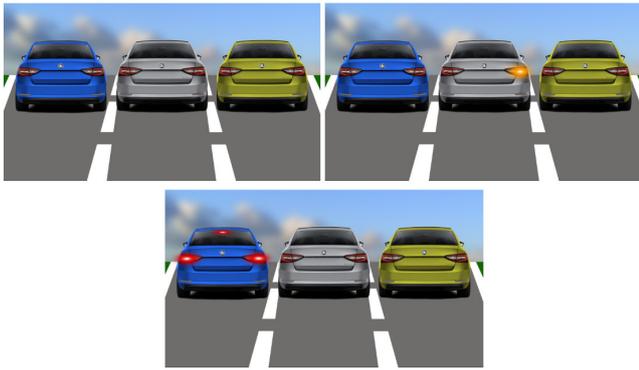


Fig. 1. Screens of primary task script in three different states (none action, middle car signals right turn, left car is braking)

#### IV. FIRST EXPERIMENT

##### A. Participants

Target group for our experiment were people over 60 years old. Having a currently valid driving licence was not required but the participants should have driving experience. We recruited 7 participants. They were aged from 61 to 74 years ( $mean = 68.71$ ,  $SD = 5.47$ ). All of the participants were native *Anonymized* speakers. For more information see Table VI

##### B. Apparatus

**Primary task simulation.** The primary task was designed for simulation of paying attention while driving. It was being handled by a script, which showed static pictures of three cars (see in Figure 1) from the back view on the computer screen. After a period between 6 to 10 seconds (randomly selected), one of the cars performs action of braking, signals turning right or left (illustrated by turning on back brake lights or blinking of signal lights). All of these parameters are chosen randomly. Participant should react immediately to these actions by pressing the correct key on the keyboard (spacebar - brake lights, left/right arrow for turning signals).

**Secondary task.** Wizard of Oz technique was used for simulation of interaction with designed dialogue system for each task. For this experiment moderator played pre-recorded phrases according to the participant's responses and the current state of dialogue. For the uncovered states (unexpected participant's answer or request), universal recovery phrases were prepared. Furthermore, the participant could ask for repetition of question/answer. Playing of pre-recorded phrases was handled by the moderator using the web application (HTML and JS), which allows controlling playback of phrases for specific states. The control interface also contains a dialog state diagram (see example in Fig.2) for the given scenario for better orientation of moderator in the dialog flow.

**Equipment.** The experiment was done by using two notebooks, the primary task simulation script was running on the first one (with an external keyboard connected) and the second one for controlling the Wizard dialog application by

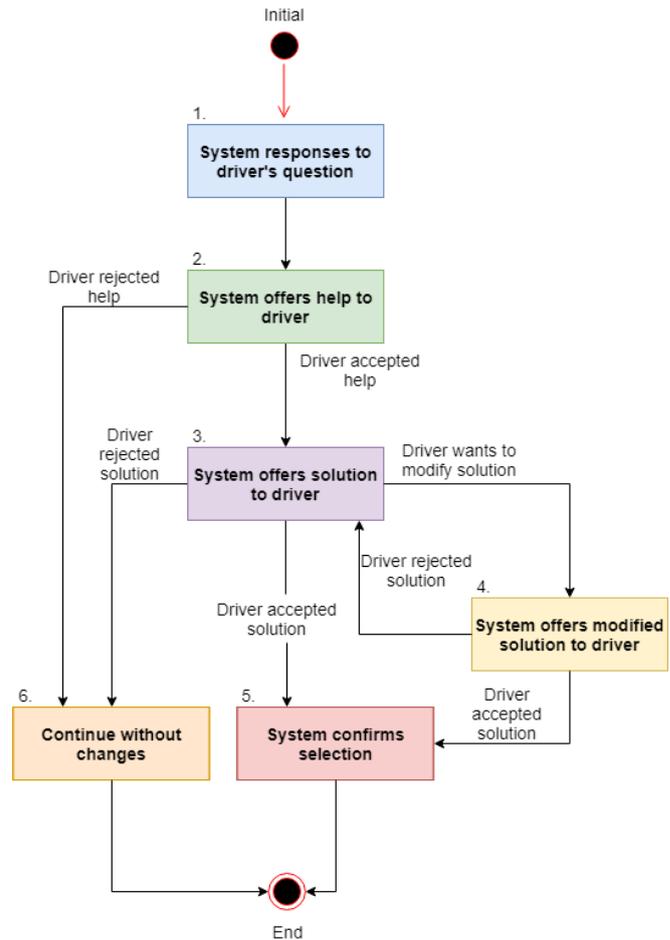


Fig. 2. Dialog state diagram example (Task B1).

the moderator. Each use-case was complemented by a printed map for better illustration of situational context described in the scenario presented to the user. Furthermore, we used a small car model for showing participant's position on the map. Experiment setup (without the moderator's notebook) can be seen in Figure 3.

**Data Collection.** During each session, the audio was recorded. The reaction time and correctness of the pressed key were also measured. However, the number of participants is not sufficient for quantitative testing and measured values are not significant for this time, it can be used for later experiments or evaluation.

##### C. Procedure

The experiment was divided into the following individual tasks, which were based on previously described navigation and messaging tasks with selected scenarios.

###### 1) Experimental tasks:

- Training: primary task only
- **A1**: Stressful situation on the crossroads - false negative
- **A2**: Stressful situation on the crossroads - false positive
- **B1**: Low fuel

TABLE VI  
TABLE WITH DETAILED INFORMATION ABOUT PARTICIPANTS

Participant ID	Balancing	Age	Gender	Active driver	Infotainment exp.	Voice interface exp.	Visual Impairment
P01	AB	61	male	yes	yes	yes	reading and distance eyeglasses (5-6 dioptre)
P02	BA	62	female	yes	yes (not using navigation)	no	distance eyeglasses (2 dioptre)
P03	AB	74	male	no	no	no	uses eyeglasses only when the light conditions are not good
P04	BA	74	male	yes	no	no	reading eyeglasses
P05	AB	67	male	yes	yes	no	no
P06	BA	72	male	yes	yes	no	distance eyeglasses (2 dioptre)
P07	AB	71	female	yes	no	no	uses eyeglasses for reading and driving



Fig. 3. Experiment setup, without second notebook for experienter

• **B2:** Semi-autonomous messaging assistant

2) *Balancing*: For the experiment we chose the balancing of the experimental tasks as AB – BA. With that every participant will go through Training phase, interacting only with primary task. Task assignment to participants can be seen in the Table VI.

3) *Surrounding scenarios*: Before the beginning of each task, it was necessary to empathize the participant into the situation. For that we used following surrounding scenarios together with printed maps (map for task A1 and A2 can be seen in the Figure 4, map for tasks B1 and B2 in the Figure 5). After the simulation is started, the moderator initiates interaction with the participant through a dialog system, following the state diagram of the dialog (this is repeated for all tasks).

**A1.** The driver arrives at a crossroad in the right turn lane. The moderator shows to the participant where his/her final destination is, using the map, and where it is best to get to it (turn left). At the same time, moderator points out that the participant is in the right lane and on the traffic situation in other lanes. The participant begins with an imaginary approach

to the intersection and the moderator launches the primary task simulation.

**A2.** The driver arrives at a crossroad in the right turn lane. The moderator shows to the participant where his/her destination is, using the map, and where it is best to get to it (turn left). At the same time, moderator points out that the participant is in the right lane and on the traffic situation in other lanes. However, the participant is instructed to "make a small break" and stop at a fast food restaurant that is a few meters after turning right. The participant begins with an imaginary approach to the intersection and the moderator launches the primary task simulation.

**B1.** The driver is on his/her way home. Finding out that he/she does not have too much fuel and many miles ahead of him/her, theoretically, there might not be enough fuel in the tank to complete the trip. The participant is also informed that s/he owns a Shell VIP card, thanks to that a liter of gasoline is much cheaper than the normal price. The participant is instructed to try to deal with a possible fuel shortage. To add more weight to a given situation, s/he is once again informed of the long journey waiting for him/her and the fuel tank indicators, which for the time being are not signaled by the indicator light, but it is clear that this may happen soon. The participant starts when s/he is already on the road and the moderator launches the primary task simulation.

**B2.** The driver sets out on his/her way home to his son David. S/he is very caring and would like to have an overview of the course of the driver's journey (when he set off, delay, etc.). The moderator informs the participant that it is possible to activate the messaging assistant at the beginning of the trip, which can help him/her to inform the son. It is up to the participants to activate the assistant. Moderator informs the participant that use-case ends after imaginary arrival home. The participant begins the situation when he sits in the vehicle and sets out on the road. Moderator launches primary task simulation. After the simulation is started, the moderator waits for the participant's stimulus and then initiates interaction with the dialog system by following the status diagram of the dialog. If the suggestion does not come for a longer period of time, the moderator initiates the dialogue initiation phrase (in the post-questionnaire asks why the participant did not start the communication).

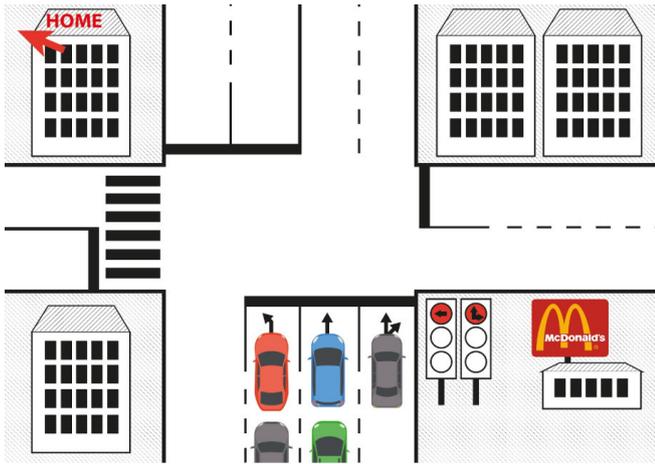


Fig. 4. Map for better illustration of the Navigation scenario

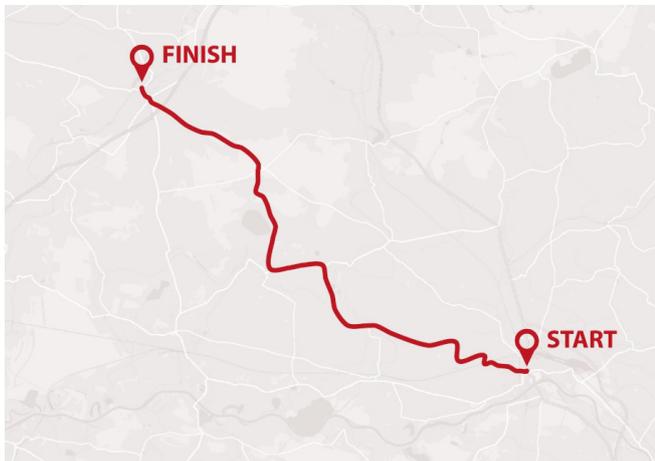


Fig. 5. Map for better illustration of the Navigation scenario and Messaging scenario

4) *Post-interviews*: Three post-interviews were executed to get feedback from participants about the subject of testing. One cumulative for tasks A1 and A2, and two separate for each B1 and B2 task. The open questions about their first impressions and how did they liked the dialogue with the conversational assistant, were posed to participants. We also asked participants about their subjective judgements about the level of comfort (“*I felt comfortable when I was using the system.*”), comprehension (“*I think that the system spoke comprehensively.*”), intuitive conversation (“*I think that the conversation was intuitive.*”) and level of acceptance in traffic (“*I can imagine that I would use the system in traffic.*”) on a 5 point Likert scale as a level of agreeing with presented statements.

## V. RESULTS

### A. Wrong lane on the crossroad

**Trust and distraction.** In general, situation in scenario A was hard to imagine for participants, P07 said she had never

been in this situation in the real life, P02 did not trust the advice from the system to continue straight, because it is not sure what will be the situation on the next junction, P03 cannot imagine using the system in this situation. He does not like larger dialogues while driving (also with passengers), he rather focus on driving. P06 said it was excellent, that system reacted immediately to the situation. Regarding the deviation to the restaurant, it understood me and provided me even with the distance information.

**Complexity of answers.** Participants P01, P02 preferred shorter speech. They answered mostly with short answers, he used answers like: “*Yes/Yep/No*”, etc. , therefore, did not get enough information to go into the right state. On the other hand, P06 answered the first question with a complex utterance (“*Yes, but I will change the route to the McDonald’s restaurant*”).

**Modification of system suggestion.** Some of the participant modified the suggestion from the system, P01 for instance, did not declined or accepted the suggestion from the system, but said: “*To the right*”. Even if he was told he want to turn left to reach the destination, he chooses the third option. P05 replied yes to the question if he cannot turn left (“*It seems that is hard go left now, is this true?*”). But the answer to the question if he wants to continue straight was: “*Not really, we will try it for now.*”

P04 would like to have more additional information for making his decisions. For instance availability of parking lots, which can be also found on road signs, but they are not so important and he can miss them because he is focusing on the driving.

### B. Low fuel situation

**Start of the conversation.** P01, P03, P04, P07 did not initiate the dialogue by themselves or were not sure how to initiate it. P04 expected that the system will offer him his preferred petrol station automatically. P05 was continuously talking about the situation (thinking aloud), about worries if there is enough fuel to reach the destination, it would be hard for the system to recognize his intent.

**Modification of system suggestion.** P01, P05 correctly rejected OMW station as they were instructed they had VIP card for different petrol station. P03 did not ask about searching for his preferred petrol station, he waited for information given by the system. P06 was asking the system to find Shell gas station (“*Find closest Shell*”). Participant refused offered Shell (because of the deviation from the route) and asked for another Shell gas station. System offers OMW station which is directly on the route. Participant refuses and asked for Shell station (“*How far will be Shell*”). System offered the same Shell (with deviation). Participant refused and asked for another Shell. System refused and participant asked why it is not possible to find another Shell Station. Participant tries to refine the request by specifying to find Shell station without deviation (“*Find closest Shell without deviation*”). Finally participant accepts the Shell station with deviation. P06 disliked the offer of the gas stations. There for sure must be another Shell station

closer to the route, what is a problem of the database of POIs rather than problem of the dialogue.

**General.** P01 was often answering before the utterance from system was completed, he seemed to be impatient. P02 used “Thank you” to end the conversation, she said it was great and she liked it, also understood it well. P03 did not like risky situation, when he was not sure if the amount of petrol is enough to get to final destination. P04 did not know that he can ask about his favorite petrol station, so the system was still repeating information about the same petrol station. It was annoying for him. On the other hand, P07 liked the fact that she could ask the system about different petrol stations.

### C. Messaging assistant

**Distraction.** P01 mentioned that the sending of messages should be automatic, when he starts the route navigation. P01 also said the amount of the conversation during this scenario was too big. He would like it to be more automatic, because this semi-automatic system can be still too distracting. P03 appreciated the ability to send message without distraction.

**Message preparation.** Participant P05 dictated all parameters in one sentence (time, contact phone number, message). P06 was dictating the content of the message only and as an appendix adds “...and then I will finish the sentence”.

**Sending message.** On the way P06 asked to send a message to David by means of clear request to the system (“Send a message to David.”). Participant requested to add time to arrive to the message.

**Complexity of answers.** P05: The answer to the question “Do you want to send an SMS” was complex “Well, definitely, because I will be late, I guess.”. The answer to the question “Do you want to add something to the SMS” was also complex “No, that is enough, it is exactly what I meant.”

**General.** P02 started conversation or activated the messaging system with “I’m just leaving”, system then asked about the purpose of the drive, participant answered “I am going to see you”. Participant P03 appreciated using of messaging assistant. P06 requested confirmation of the message delivery.

### D. Post-interview

P01 said that the system was too verbose. P02 mentioned it would be difficult for her to learn how to use this system, because she does not use the navigation system in her car and she drives only on known routes. On the other hand, she would welcome the messaging system in her car. P03 had doubts about the reliability of these systems in nowadays cars in general. P05 mentioned he was a bit nervous about what and when will happen, it means it is not clear when and what the system will start talking about. “The system could maybe continuously talk about the situation. Long silences makes me nervous if the system is functioning and the question is surprising me.” P07 liked that he can control message sending with voice and did not have to use hands or make a phone call.

### E. Subjective judgements

Fig. 6 shows that system acceptance in traffic was positive mainly for tasks B1 and B2, regarding tasks A1, A2 almost half of the participants would not accept this system in real traffic situations. Almost 100% of participants strongly agreed that the conversation with system was intuitive, only one participant was neutral in task B2. 90% of participants agreed or strongly agreed that the system spoke comprehensively. 2 participants disagreed in task B2 about the comprehension. 95% of participants agreed or strongly agreed the system was comfortable to use.

## VI. DISCUSSION

The qualitative results show that drivers often did not know how to start a conversation with the system. First, older drivers have mostly no voice control experience. Second, the tasks were hard to imagine for the participants. It was our intention not to tell them how they should start the conversation before we started the study, and we just wanted to observe their behavior without previous experience with our system.

The complexity of the participant’s utterances to the system varied widely across the participants, some were talking to the system in complex sentences so that current natural language understanding systems would have a problem identifying the user’s intent. Conversely, there were users who talked to the system very briefly and austere.

The messaging assistant was perceived with a predominantly positive attitude among the participants. They would find it practical and could imagine using it in the car. Some participants would expect the messaging assistant to be even more autonomous and found it unnecessary verbose. Here arises an opportunity to personalize the assistant, for example, based on verbosity and level of automation.

In the navigation task, the biggest problem for the participants was to imagine the scenario situations, but many of them were positive about the contextual information the assistant offered them. The results show that the assistant should be able to offer alternatives, for example, when choosing a gas station. For instance, an assistant informed in advance about the low fuel level, but with the assurance that there is still enough to reach the destination, the participants welcomed and then freely decided whether to stop at the gas station or not. Similarly, in the simulated crossroad situations, drivers freely decided whether to accept the system’s recommendations or not.

### A. Limitations

The main limitation of a user study we conducted was the fact that we did not use a high-fidelity driving simulator instead, we used our low-fidelity simulation of the primary task described in the section IV-B. Therefore, some of the participants had a problem to imagine the context in scenarios we presented them. Furthermore, because of our experimental setup, we were not able to simulate real stressful road situations and so we do not claim that our design will reduce the stress of older drivers, but rather see an opportunity

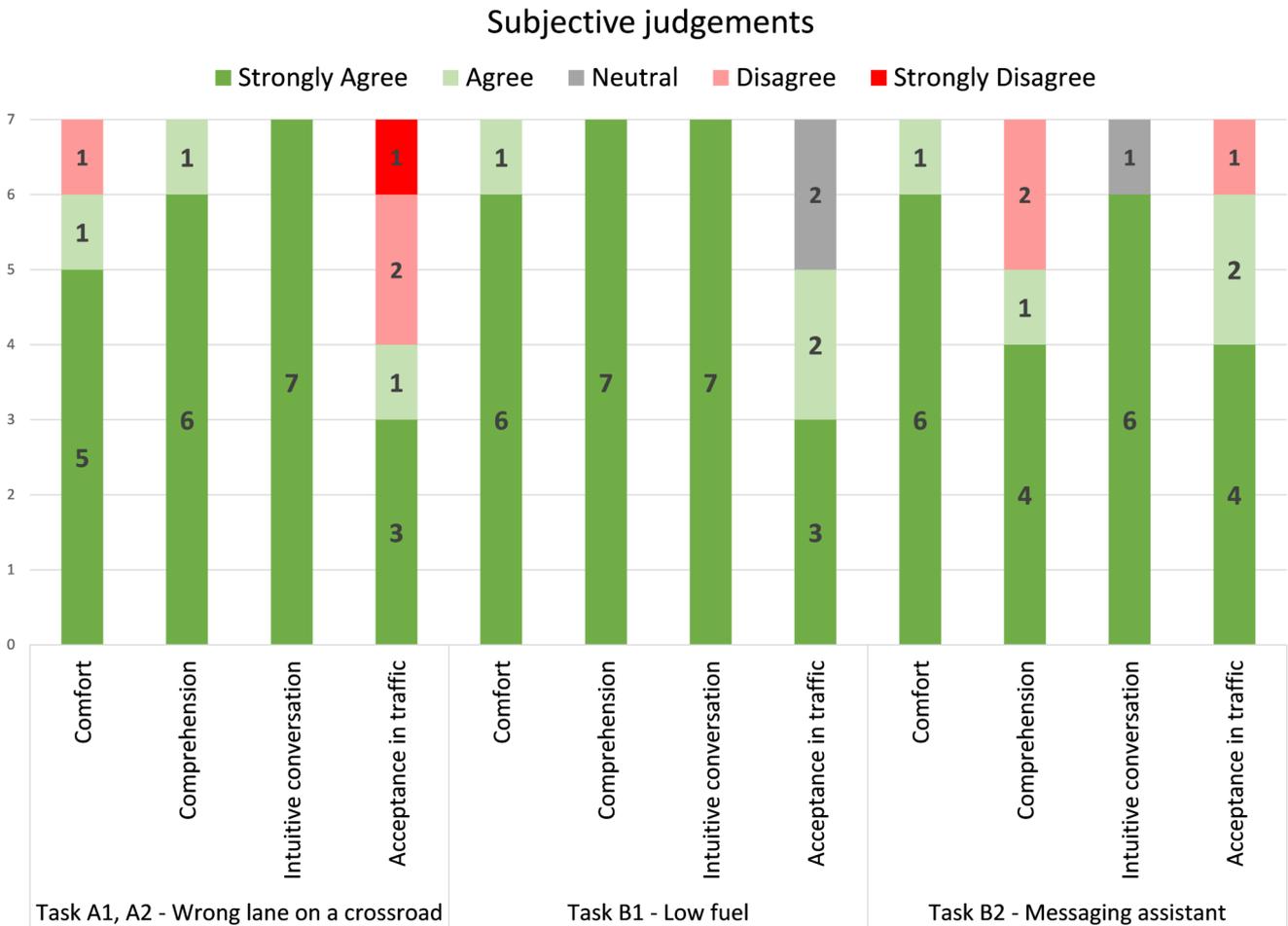


Fig. 6. Subjective judgements about level of comfort, comprehension, intuitive conversation and acceptance in traffic ( $N = 7$ ).

in development such intelligent system which would respect older drivers according to our findings.

### B. Future Work

Given our findings and limitations, future work should further investigate the use of other modalities for supporting the older drivers in tasks that require more rapid action from the driver. For instance in situations on the crossroad and also other fast and stressful situations, the haptic interface, we believe, could be more suitable.

We would also like to further explore the ways how to announce the driver that system is going to talk to him/her, as some participants were surprised when the system started to speak after a longer pause. Again multiple modalities should be investigated. The use of the context of the car, road situation, surrounding traffic or drivers preferences and skills should be explored to determine when is the right time for the system to start a conversation with the driver.

## VII. CONCLUSION

We designed a low-fidelity prototype of conversational assistant for two secondary tasks, navigation and messaging. We also conducted a qualitative user study with 7 older drivers, using the Wizard of Oz method. The results of our study show that support in secondary tasks for older drivers while driving can be carried out by the conversational assistant. But still, there are some limitations when using only speech-based interface, therefore, the use of other modalities should be investigated.

### ACKNOWLEDGMENT

This research has been supported by grant no. SGS19/178/OHK3/3T/13 (FIS 13139/161/1611937C000) and by project RCI (reg. no. CZ.02.1.01/0.0/0.0/16 019/0000765) supported by EU.

### REFERENCES

- [1] Czech Statistical Office, "Seniori," [Online; accessed 13-May-2019]. [Online]. Available: <https://www.czso.cz/csu/czso/seniori>

- [2] BESIP, "Senior v silnicnim provozu," [Online; accessed 13-May-2019]. [Online]. Available: <https://www.ibesip.cz/Tematicke-stranky/Seniori/Senior-v-silnicnim-provozu>
- [3] H. C. Lee, A. H. Lee, D. Cameron, and C. Li-Tsang, "Using a driving simulator to identify older drivers at inflated risk of motor vehicle crashes," *Journal of Safety Research*, vol. 34, no. 4, pp. 453 – 459, 2003, senior Transportation Safety and Mobility. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0022437503000471>
- [4] S. G. Klauer, T. A. Dingus, V. L. Neale, J. D. Sudweeks, D. J. Ramsey *et al.*, "The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data," 2006.
- [5] E. Chan, A. Pradhan, M. Knodler, A. Pollatsek, and D. Fisher, "Evaluation on a driving simulator of the effect of drivers' eye behaviors from distractions inside and outside the vehicle," *Human Factors*, 2008.
- [6] J. Fofanova and M. Vollrath, "Distraction while driving: The case of older drivers," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 14, no. 6, pp. 638 – 648, 2011, special Issue: Driving Simulation in Traffic Psychology. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1369847811000775>
- [7] D. Shinar, N. Tractinsky, and R. Compton, "Effects of practice, age, and task demands, on interference from a phone task while driving," *Accident Analysis & Prevention*, vol. 37, no. 2, pp. 315 – 326, 2005. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S000145750400096X>
- [8] M.-P. Bruyas, C. Brusque, S. Debailleux, M. Duraz, and I. Aillerie, "Does making a conversation asynchronous reduce the negative impact of phone call on driving?" *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 12, no. 1, pp. 12 – 20, 2009. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1369847808000594>
- [9] J. Y. Lee, M. C. Gibson, and J. D. Lee, "Error recovery in multitasking while driving," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ser. CHI '16. New York, NY, USA: ACM, 2016, pp. 5104–5113. [Online]. Available: <http://doi.acm.org/10.1145/2858036.2858238>
- [10] J. Maciej and M. Vollrath, "Comparison of manual vs. speech-based interaction with in-vehicle information systems," *Accident Analysis & Prevention*, vol. 41, no. 5, pp. 924 – 930, 2009. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0001457509001080>
- [11] B. Pfleging, S. Schneegass, and A. Schmidt, "Multimodal interaction in the car: Combining speech and gestures on the steering wheel," in *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, ser. AutomotiveUI '12. New York, NY, USA: ACM, 2012, pp. 155–162. [Online]. Available: <http://doi.acm.org/10.1145/2390256.2390282>
- [12] G. Shakeri, J. H. Williamson, and S. Brewster, "Novel multimodal feedback techniques for in-car mid-air gesture interaction," in *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, ser. AutomotiveUI '17. New York, NY, USA: ACM, 2017, pp. 84–93. [Online]. Available: <http://doi.acm.org/10.1145/3122986.3123011>
- [13] I. Tashev and Y. C. Ju, "Commute ux: Voice enabled in-car infotainment system." Association for Computing Machinery, Inc., September 2009. [Online]. Available: <https://www.microsoft.com/en-us/research/publication/commute-ux-voice-enabled-in-car-infotainment-system/>
- [14] A. Bjelkemyr, T. Dukic, R. Owens, T. Falkner, and H. Lee, "Support systems designed for older drivers to achieve safe and comfortable driving," *Journal of Transportation Technologies*, vol. 3, pp. 233–240, 2013.
- [15] A. Ziakopoulos, A. Theofilatos, E. Papadimitriou, and G. Yannis, "A meta-analysis of the impacts of operating in-vehicle information systems on road safety," *IATSS Research*, 2019. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S038611121830030X>
- [16] K. Lipovac, M. Đerić, M. Tešić, Z. Andrić, and B. Marić, "Mobile phone use while driving-literary review," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 47, pp. 132 – 142, 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1369847817302838>
- [17] P. Atchley, S. Atwood, and A. Boulton, "The choice to text and drive in younger drivers: Behavior may shape attitude," *Accident Analysis & Prevention*, vol. 43, no. 1, pp. 134 – 142, 2011. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0001457510002095>
- [18] I. A. Al-Darrab, Z. A. Khan, and S. I. Ishrat, "An experimental study on the effect of mobile phone conversation on drivers' reaction time in braking response," *Journal of Safety Research*, vol. 40, no. 3, pp. 185 – 189, 2009. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0022437509000334>
- [19] C. S. Dula, B. A. Martin, R. T. Fox, and R. L. Leonard, "Differing types of cellular phone conversations and dangerous driving," *Accident Analysis & Prevention*, vol. 43, no. 1, pp. 187 – 193, 2011. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0001457510002241>



# Supporting personalized care of older adults with vision and cognitive impairments by user modeling

Petr Bilek, Miroslav Macik, Zdenek Mikovec

Department of Computer Graphics and Interaction

Faculty of Electrical Engineering, Czech Technical University in Prague

Karlovo nam. 13, Praha 2

Email: macikmir@fel.cvut.cz

**Abstract**—We present a user modeling approach tailored to the user group of older adults with vision impairments. Our previous work with this user group has been followed by qualitative user research with the staff of a residential care facility for visually impaired older adults. We defined the structure of a user model that represents aspects related to personal psychological development and attributes that affect interaction with technologies. Furthermore, we present two generations of prototypes of administration UI for the user model. Results of the qualitative evaluation are discussed in the paper. This effort aims to help the personnel of specialized care institution in providing personalized care and make the interaction with technologies accessible for the group of visually impaired older adults with vision impairments.

## I. INTRODUCTION

ACCORDING to [1] the prevalence of blindness is globally 0.48 %, and 2.95 % of world population deals with moderate to severe vision impairment. The majority of visually impaired people appear among older adults, as 52.9 % of visually impaired people are older than 70 years.

The proportion of older adults in the population is steadily rising. In addition, according to WHO this trend will continue (due to increasing life expectancy and decreasing birth rate) [2]. During ageing, a large proportion of the population is acquiring some kind of health impairment. This fact can be also seen in data of Czech statistical office (data were processed in 2013) [3], 18.6 % of people between the ages of 60 and 74 have some kind of disability. In category 75+ years old, percentage of disabled people rapidly rises, it is around 42.1 %.

Unfortunately, the research attention on older adults with vision impairments (VI) is limited. We have analyzed 39 papers focusing on people with VI presented on last three CHI conferences (2016-2018). When excluding studies focused mainly on children and young adults, the average age of study participant was 37.3 years (weighted average, sample sizes as weights). Hence, the current research focus is in the case of VI people biased in favor of the younger part of the population.

User research conducted in a specialized residential care facility for visually impaired older adults [4] showed that clients are also often challenged with other age-related health issues. Frequently, those people are challenged with cognitive issues (mainly dementia), or mobility issues. The facility

provides individualized care based on personal psychobiographical modeling of individuals [5], [6].

In our previous research, we focused on the development of solutions to support orientation in space tailored to needs and preferences of visually impaired older adults [4]. The evaluation indicated that adaptation of user interfaces and interaction method based on individual needs and preferences could improve usability of particular solutions as well as maximize their acceptance among the target user audience.

In this paper, we present a user modeling approach tailored primarily for the user group of visually impaired older adults living in residential care facilities. The user model can represent individual aspects related to personal biography as well as aspects important for adaptation of user interfaces and interaction methods.

## II. RELATED WORK

In this section, we discuss various technical approaches to represent personal attributes related to interaction with other human beings as well as with technologies – user models. Furthermore, we describe specific approaches for geriatric and gerontopsychiatric care. Finally, we list technological approaches where user model is used for adaption of user interfaces and interaction methods and for providing personalized care.

The original user modelling approaches emerged from the medicine and rehabilitation engineering. The World Health Organization (WHO) comes up with models, which are based typically on measurement and quantification of human performance, which is examined in rehabilitation engineering. The WHO defined ICD (International Statistical Classification of Diseases and Related Health Problems) [7] and ICF (International Classification of Functioning, Disability and Health) [8].

Ability-based design method proposed by Wobbrock et al. [9] suggests leveraging specific individual abilities rather than focusing on disabilities. The use of user-specific abilities and capability of interactive systems to adapt accordingly could make the interaction more efficient, natural, and broaden the audience of users that can successfully use a particular technology.

Peißner et al. [10] come with the idea of individual patterns for accessible and adaptive user interfaces, which are built on information about the user, the context and the devices,

which is gathered through system interaction and sensors. All information about the user is stored in the User profile, which variables include his abilities, disabilities, preferences and his current environment. All the values of the user profile are numeric between zero and four. The context manager continuously updates the user profile. The user profile mixes different variables, and some of them are not so typical for user models (e.g., ambient light and ambient noise).

Heckmann et al. [11] introduce the general user model ontology (*GUMO*) for the uniform interpretation of distributed user models in intelligent semantic web enriched environments. *GUMO* uses *UserML* (User Model Markup Language) [12], which is Resource Description Framework (RDF)-based user model exchange language.

Kikiras et al. [13] present a user model for navigation systems, which is represented through a Semantic Web ontology (User Navigation Ontology – *UNO*). For that purpose, they use the Web Ontology Language (OWL) for describing the user classes and their properties. During ontology development, they extended some of the concepts specified in the *GUMO* ontology [11]. They also adopted the International Classification of Functioning, Disability and Health (ICF) [8] of World Health Organization (WHO) for representing certain functioning and disability issues of an individual. Their model comes up with so-called "*User profile*", which captures user's demographics, mental-cognitive characteristics, sensory abilities, motor capabilities, navigational preferences and interface preferences.

The Böhm's model [5], [6] is an internationally recognized nursing model which is currently used at most in German-speaking countries in the field of geriatric and gerontopsychiatric care. The model is aimed to support the self-care of old and confused people and it is also focused on how to retain or restore the self-care ability for as long as possible by the principle of recovery of the senior's interest by reviving his psyche. Daily life and normality for patients is one of the core issues of Böhm's nursing model. The theory of Böhm is based on understanding of patients, their biography and coping strategies. The biographic assessment and knowledge of coping strategies is crucial for the psychobiographic nursing concept. Patients should stay as long as possible independent and self-reliant and should keep their social competence. Böhm comes up with the theory that learning copings takes place in the first 25 years of life, and in regression, one returns to their own lower copings.

Hoey et al. [14] present a summary of customizable and adaptive technologies for assistance for persons with cognitive disabilities. Authors describe decision-theoretic model based on Partially Observable Markov Decision Process (POMDP) that can be applied to various activities. The system allows user customization, system adaptivity to user and general purpose sensing abilities. Authors state that the concept of *inclusive design* works only for large and uniform segments of population, but fails for individuals with diverse and changing needs like older adults with Alzheimer's disease. A system itself must adapt to people over time as they change.

Internet of Things sensor network as source of heterogeneous contextual information is mentioned in [15]. The contextual information (e.g. daily activities) is integrated into personalized care management processes to support automatic and better decision making in an effective and user-friendly manner. There is set of sensors that recognize personal activities, and the system can provide guidelines how to proceed when one experience an issue when carrying out a particular task. These context-aware services can help older adults to stay at their homes safely.

There are several approaches for user modeling, however, only few leverage specific user abilities (what user can do and can be used for interaction) rather than representing disabilities (what user cannot do and must be compensated by assistive technologies). User group of visually impaired older adults requires a specialized approach for both care provided and interaction with technologies. User research described in section III leverages these specifics. The available user modeling approaches cannot represent the necessary user attributes to a sufficient extent. Also, there is a lack of technological support for psychobiographical modeling of individuals and corresponding care model.

### III. USER STUDY

To learn more information about application of psychobiographical care model in practice, the user research was conducted in the institution for visually impaired people (104 employees and 125 clients in 2019).

#### A. Participants

User research was conducted with three participants (P1–P3), all women, average age 44.33 ( $SD = 15.95$ ,  $MIN = 26$ ,  $MAX = 55$ ). *P1* and *P2* work as activation services workers and *P3* works as direct care worker (social worker). None of the participants worked in a similar type of facility before working in the residential care home we cooperate with. Mean duration of current job title was 7.83 years ( $SD = 7.18$ ).

#### B. Procedure

The user research has the form of a qualitative semi-structured interviews with employees of the institution for visually impaired people. The topics were focused on the activities of activation services workers and direct care workers, the usage of the psychobiographical care method in practice, and the usage of technologies in the institution. One interview lasted approximately 35 minutes (short briefing, 30 minutes interview and short debriefing).

#### C. Results

**Job description.** The main task of the **activation services worker's** job is to give clients the opportunity of self-realization even in older age. Activation services workers prepare for clients various types of activities, both individual and group. These include, for example, cooking, singing, memory training and many more. The activities are tailored to individual clients according to their preferences and habits

from their previous life and are often related, for example, to the client's profession. All activities are voluntary, and it is up to clients whether they want to do any of them. Activation services workers cooperate with health professionals and direct care workers, and it is very important to create one compact team of workers.

**Direct care workers** are responsible for the social aspects of care. They help clients with a wide range of activities, such as personal hygiene, dressing, moving, and eating. An important part of their work is also communication with clients and supporting them in their favorite activities. Every direct care worker is in charge of three clients as a so-called "key worker". He/she should know the most about these clients, map their needs and preferences and share this information with the entire team.

**Daily programme.** The residential care home daily programme has three reference points – breakfast, lunch, and dinner. Between breakfast and lunch, and then between lunch and dinner, there are group, as well as individual, activities that are provided by activation services workers and direct care workers. For clients, the program is voluntary, and it depends only on them what activities they choose, each client has different interests and preferences. Within the individual activities, workers chat with clients, read, sing, go out, etc. There are plenty of group activities in the residential care home – rehabilitation exercises, workshops, singing, cross-words solving, memory training, art therapy, canistherapy, music therapy and more. Group activities take place regularly, usually once a week. Cultural events, such as concerts or theatre performances, are held twice a week at the Great Hall, and these events are very popular. There are also held, approximately four to five times a year, excursions outside of the residential care home, for example, trips to castles and chateaus, trips to theatres and a trip by steamboat.

**Technology use by clients.** The use of computers and similar technologies is limited in the residential care home due to various limitations of clients. Most clients also have a barrier between them and modern technologies. "People, clients, who come here can't work with modern technologies. Maybe they have never seen a computer during their lifetime." (P2)

In the residential care home, there is a device called a reading magnifier, which is used by some clients. In some departments, some gramophones are used to play old records and also to induce the atmosphere of the times when gramophones were commonly found in households. In addition, they use voice-output clocks to inform clients of the current date, time, weather, and so on. Unfortunately, for most clients, these clocks are too complicated. "The voice clocks we use here are too complicated. For most clients, clocks report too much information, which can lead to a big cognitive burden on clients." (P3)

#### **Biography and Böhm's psychobiographic method.**

In the residential care home, Böhm's psychobiographical model of care is applied, specifically its modification, which is adapted to work in the institution. "We use the Böhm's method

*adapted for our needs.*" (P3) The client's biography is a key element of the psychobiographical model. "Biography is such a client's book of life." (P2)

The process of collecting information into biography and its processing begins in the period before the client comes to the residential care home. In this period, the client and his family may be asked to provide some information. "Even before the client arrives, the client and his/her family are asked to think about providing information and photos of the client and to write a client's life story at their discretion." (P2) Unfortunately, cooperation with family and clients is sometimes difficult. "Sometimes collaboration with client's family is complicated because they are reluctant to provide sensitive information about a close person." (P2) However, it is essential that neither the client nor his/her family is pushed into anything and that the provision of information must be based on a free decision.

After the client's entry to the facility, the "key worker" assigned to the client has to start the processing of biography. "We then communicate with the client and start processing the mapping, where we write down the information that client provides to us. Based on it, we create the client's life story, which is written into a biographical book with attached photographs." (P2) Workers must establish a relationship with the client to get as much information about the client as possible. "The amount of information that clients are willing to convey is different, someone is willing to tell everything, someone almost nothing. It is up to us to establish a relationship with the client to tell us something, but at the same time, we must understand that client have told the information to us and may not want the information to appear somewhere. I had a client who directly wished for some information not to appear anywhere and did not even want to publish information under her name, so she invented a fictitious one." (P3)

Biography is in paper form. It has a fixed structure according to which workers must proceed during its creation process.

In the biography, there is a client's life story with photographic documentation. The life story contains information about the client's childhood, youth, adulthood, old age, and this information is then used to work with the client. There is a need to find out what the clients have gone through in their lives, about their families, their interests, their education, their occupations and then specific patterns of behaviour can be derived from this knowledge. "Biography should tell us about client's behaviour patterns, in which environment he grew up, and we must realise that man in old age is hardly going to change his behaviour." (P3) It is also necessary to put the information about the client's life in the historical context. Many clients have experienced the First Czechoslovak Republic, the Nazi occupation, the putsch in 1948, the totalitarian regime, the Soviet invasion in 1968, and of course all these events influence man. "We should have a basic, maybe even advanced, knowledge of history to know what it does to a man when he experiences during his active life such turbulence as in the 20th century in Europe, especially in Czechoslovakia." (P3) It is also necessary to learn negative information about

the client to avoid possible misunderstandings and unpleasant situations. *“It is also good to know topics that client doesn’t want to talk about. Some clients have experienced war or totalitarianism. Some of them don’t mind talking about it, but some clients don’t want to hear German or Russian. So it is necessary to know negative information, experiences, etc.” (P2)*

Also, the biography includes a client’s tree of life (lineage), where information about his relatives, friends, etc. can be found. *“In deeper dementia, clients become more likely to think that we, caregivers, are their friends when they were ten years old and they address us with the names of their friends. And when we know that he/she is addressing us on behalf of a friend, we know he/she has such a relationship with us.” (P2)*

Furthermore, so-called *“activities of daily living”* are part of the biography. These are divided according to the tasks of the ordinary day. They describe what is typical for the client, what he was used to and how to transfer this fact to residential care home reality as faithfully as possible so that the client feels at home. The biography also affects activities such as client dressing up, which can be taken as an example of such a task of daily living. *“We had a client who belonged to the almost highest social class in the First Czechoslovak Republic. She was very fond of wearing luxury costumes with pearls. Thus, in dressing, it would be written in the activities of daily living, that she was used to the luxury, that she still keeps these habits and that even when she is infirm, we will not give her sweatpants even if it is easier for us to provide her with care.” (P3)*

If a client with dementia enters the facility, it is necessary to collect the information as soon as possible. *“When a new client with dementia comes here, we need to find out as much information as soon as possible. There is a need for cooperation with the client’s family. It may happen that the client enters the residential care home and in two months period, he will not be able to provide us with further information. Then, work based on the biography is complicated.” (P2)*

Biography is constantly maintained and updated with new information. *“Biography is handled by everyone except medical staff. Maintenance of biography is the work of the whole team.” (P2)* Access to biography, which is located in the nurses’ room in locked safes, is available to all department workers and can be provided to the client or his/her family upon the client’s request. *“Biography contains sensitive personal information, often even intimate, so trainees or anyone else who does not work in the department does not have access to it.” (P2)* Employees have regular weekly meetings where they discuss individual clients. *“We always have a meeting on Wednesday, where clients are discussed according to Böhm’s method. We all, including health professionals and a psychologist, discuss the client’s score (differential diagnosis), the activities of daily living and so on. The key worker then selects all the relevant information he needs for further working with the client.” (P2)*

The biography is fundamentally reflected in the work of caregivers. *“Activation and communication with the client are tailored to the biography.” (P2)* *“For example, the client was a seamstress all her life, so she works with a fabric that reminds her of a long period of her life.” (P1)*

Based on a biography, caregivers try to bring clients closer to the environment they have been familiar with during their active life. For example, listening to music and recordings from that period can help. *“The music of the 1950s is playing right next door – Chladil, Simonová and other singers. We play recordings from Semafor every day, so these days we get slowly to 1960s.” (P2)* Then, for example, reading books about the time, when clients actively lived, about places, which clients could visit and about the activities they could perform can also help to induce well-being. *“We read autobiographies of actors, musicians, and writers. We also read about countries and places where clients often travelled in the 1970s – Hungary, Poland, East Germany, we read for example, about Rügen and Lake Balaton. We also read about camping or pig-slaughtering.” (P2)* The aim is to retrospectively return clients to a time when they were actively living their lives. *“We are going back to the age when clients were about 25 years old.” (P2)*

For less oriented clients, workers read fairy tales, various stories and poems. *“We read mainly the fairy tales of Božena Němcová, Karolína Světlá and other authors. We read especially the shorter ones, such as The Red Riding Hood. We also read some poems that are clients familiar to.” (P2)*

The activities of clients and activities of workers with clients are recorded in the system. However, clients have the freedom of movement around the residential care home, and so, for example, a visit from a client’s family is not recorded. *“Clients have freedom of movement, but most of them report such activities. If the client leaves the home, we write down the fact that the client left and also the expected return.” (P2)*

*“The biography must be handled carefully, and workers must maintain a professional approach without an evaluation approach, even though some client’s habits may, for example, seem weird to them.” (P3)* *“The problem is with the increasing difference between clients and workers. There are two completely different worlds, the world of young workers on one side and on the other side the world of clients who are still ageing. Therefore, the worker must understand the client, which is also very difficult and some misunderstandings may arise.” (P3)*

**Technology use by workers.** A barcode readers system for recording client’s activities, which is used in the residential care home, helps employees. Each activity is assigned a unique barcode that is scanned with the reader and all records are downloaded to the system at the end of every day.

All participants can imagine the electronic version of the biography, but the current form is more acceptable and enjoy-

able for most of them (P2 and P3). “It’s nice to have it in the form of the book, it’s more natural and easier to pass on to someone.” (P2)

The user research with employees of the residential care home indicated that personal psychobiography is frequently used for purposes of providing care. The collection of data begins even before a client arrives at the institution. It contains information about client life in the historical context, their specific patterns of behavior, or activities such as client dressing customs. Furthermore, the biography contains so-called “activities of daily living” that describe what is typical for the client and what he was used to. The psychobiographical sheet, which includes all the relevant information about the client’s biography and nowadays is in the paper form, is shared among personal, and there is limited access for others. Workers also routinely record activities performed with clients into the system using a barcode scanner. The research with employees and our previous research with clients described in [4] showed, that clients have a barrier in using modern technologies. However, our previous research showed that technological products that are well-adapted to needs, preferences and abilities of them could help clients with daily activities like orientation [?] or leisure [16].

#### IV. PROPOSED SOLUTION

The whole solution consists of the user model, user interface for its maintenance and appropriate API, which will provide all the relevant information from the user model to user interface and client’s devices. The architectural overview is depicted in Figure 1. The user model and user interface are described in detail in following sections. The solution also takes into account the connection to an existing system from which records of client activities could be obtained (see Section III). Various devices will be connected to the solution. There are two groups of users, whom the solution will serve – primary group, represented by the workers (caregivers) of the residential care home and secondary group, represented by the clients, who will use the user model indirectly through their devices and provided care.

##### A. Proposed User Model

Based on the analysis and requirements of other projects, there can be deduced information categories – user model components:

1) *Demographics*: This category contains basic information about client including the client’s name, surname, birth date, birthplace, address, gender, education, department, room number, key worker, contact persons and photographs of the client. The user model also allows to model and maintain the relationship between client and workers. The relationship also captures the rights of workers (view/edit various information categories).

2) *Biography*: The client’s biography is divided into four sub-categories. The first one is the biographical sheet, which represents the life story of the client. Then, there are the

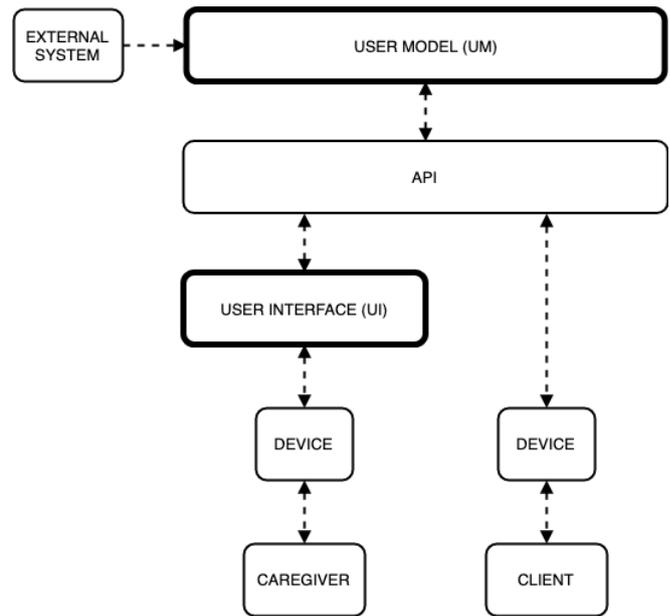


Fig. 1. Proposed solution – High level view.

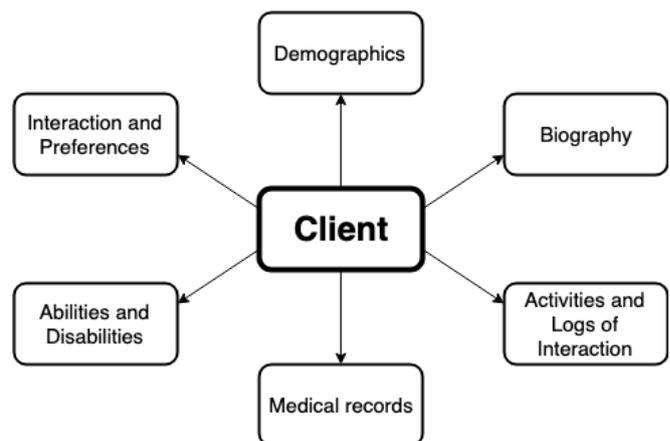


Fig. 2. User model components.

activities of daily living, which are represented by seventeen different tasks of the ordinary day (see Section III). The next category is the lineage which captures the family tree of the client. And finally the form of care, which is composed of three different types of care (activation, reactivation and stimulation) and the differential diagnostics score (eight areas of older adult behavior – psychomotor, orientation, emotions, memory, contact ability, formal thinking, will and content thinking).

3) *Activities*: This category contains records of the client’s activities. According to the conducted user research (see Section III), the employees use a barcode readers system for recording client’s activities. Each activity is assigned a unique barcode that is scanned with the reader and all records are downloaded to the external system at the end of every day. There are also logged the activity of the clients during interaction with devices, which will bring the possibility of

device adaptation.

4) *Medical records*: To this category belongs medical records of the client including the client's height and weight, diseases, injuries, allergies, diets and prescribed medication.

5) *Abilities and Disabilities*: The abilities and disabilities of the client are divided into three sub-categories – sensory, motor and cognitive.

The *sensory* category contains information about sensory abilities and disabilities of the client. Information is divided into sight, hearing and touch, additionally for each eye, ear and hand separately.

The *motor* category captures mobility information, such as motor impairment but also, for example, the ability to stand alone or ability to move independently.

The *cognitive* category contains information about cognitive abilities and limitations of the client, for example, the client's knowledge of his room or level of independence.

Information about impairment captures the severity, origin, progression and other characteristics of the impairment. There are also captures of the abilities of the client, for example, the ability to perceive object shapes or level of client's tactile abilities. If the attribute is expressed in scale, a range of 0 to 6 is used. For example for the visual impairment, the *ICD 9D90* classification [17] is used.

6) *Interaction and Preferences*: This category contains information about interaction with client and preferences of the client (for example client's dominant hand or dominant modality), including the client's attended activities, hobbies, favourite places, interests in services and client's equipment.

The categories of the user model are depicted in Figure 2.

## B. Administration UI Design

For the purpose of the creation of the prototypes, two different development tools were used. Each prototype has been implemented to support the interaction needed to accomplish the tasks performed during prototype evaluation.

1) *Low-fidelity prototype*: A very first prototype (low-fidelity prototype) of the user interface was created in the form of a paper prototype and was implemented using tool Balsamiq Mockups [18]. All application screens and all necessary interaction elements were implemented. The prototype was then evaluated in paper form (see Section V-A).

The structure of the proposed user interface reflects the user model structure (see Section IV-A) and user interface requirements, collected in time before the creation of the low-fidelity prototype. The goal was to create a prototype that is easy to maintain and minimalist, clean and intelligible. This prototype was evaluated in the usability study with three representatives of the target group. The user interface consists of five main screens, which further branches into sub-screens. The complete structure of screens can be seen in Figure 3.

**Login page.** The *Login page* is the first screen the user encounters when interacting with the application. The screen allows the user to enter username and password to log into the system.

**Home page.** The *Home page* acts as the main application signpost. It allows the user to go to the client search screen or to add a client screen. At the top of the screen, there is a panel that informs the user about the current location, allows the user to view notifications, displays the user's name and surname, and allows the user to log out of the system. This panel then appears on all other screens.

**List of clients page.** The *List of clients page* displays the list of clients and allows the user to apply filtration of records.

**Client detail page.** The *Client detail page* consists of seven sub-pages:

- **Basic information.**

The *Basic information* screen displays basic information about client. As an example of the low-fidelity prototype, screen can be seen in Figure 4.

- **Biography**

The *Biography* consists of four sub-pages:

- **Biography – Biographical sheet.**

The *Biography – Biographical sheet screen* displays biographical sheet text and attached photographs. The user is able to edit text and maintain attached photographs.

- **Biography – Activities of daily living.**

The *Biography – Activities of daily living* screen displays table with seventeen different activities of daily living. The user can edit each table row (activity) separately.

- **Biography – Lineage.**

The *Biography – Lineage* screen displays the family tree of the client. The user can maintain the lineage.

- **Biography – Differential diagnostics sheet.**

The *Biography – Differential diagnostics sheet* screen displays table with differential diagnostics score of the client and corresponding form of care.

- **Activities.**

The *Activities* screen displays activities of client. The screen is only used to show activities, the data is located in an external system (see Section IV). The user can filter activity records.

- **Medical records.**

The *Medical records* screen shows health information. It shows the client's height and weight and further diseases, injuries, allergies, diets and prescribed medication. The user can maintain all this information.

- **Abilities and disabilities**

The *Abilities and disabilities* consists of three sub-pages:

- **Sensory.**

The *Abilities and disabilities – Sensory* screen shows information about the client's sensory abilities and disabilities, such as severity, origin, progression and other characteristics of visual, hearing and tactile impairment and also corresponding abilities of the client, for example, ability to perceive object shapes.

- **Motor.**

The *Abilities and disabilities – Motor* screen shows

information about the client's motor capabilities, such as severity, origin, progression and other characteristics of motor impairment and also corresponding abilities, for example, ability to stand alone.

– **Cognitive.**

The *Abilities and disabilities – Cognitive* screen shows information about the client's cognitive capabilities, such as the ability to read Braille, knowledge of the room, knowledge of residential care home interior and other abilities.

• **Interaction and preferences.**

The *Interaction and preferences* screen displays information relevant to client interaction and client's preferences, such as dominant hand, primary modality etc. On the screen, there are also tables for records of client's attended activities, hobbies, equipment which client use, favourite places and interests in services. The user can maintain all this information, including adding and removing records in tables.

• **History.**

The *History* screen displays a history of client information changes. The user can filter records of changes.

**Add client page.** The *Add client page* allows the user to add a client into the system. Some fields are marked with an asterisk, indicating that they are mandatory. The user has the option to save the new record (and get on the Client detail page screen) or to cancel the process of the creation and return to Home page.

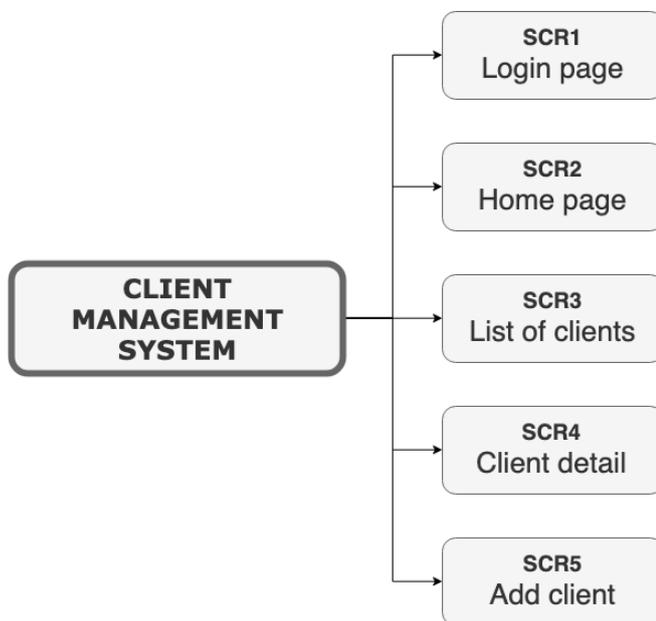


Fig. 3. Low-fidelity prototype – overview.

2) *High-fidelity prototype:* After the low-fidelity prototype evaluation, the high-fidelity prototype was implemented. The tool Axure [19], which enables the creation of more advanced prototypes, was used for this purpose. The prototype was then

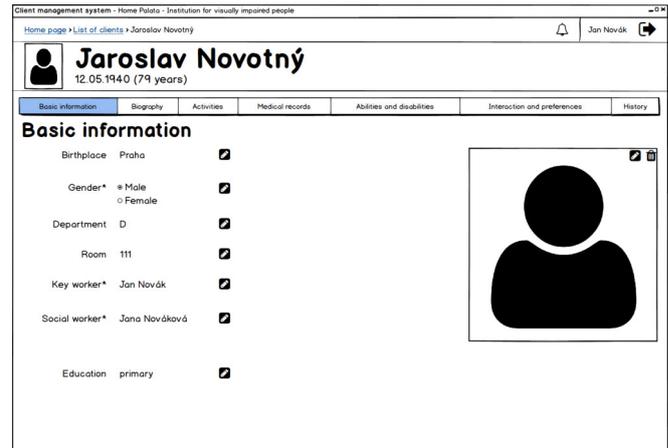


Fig. 4. Low-fidelity prototype – SCR4a – Client detail – Basic information.

evaluated in the form of interactive computer application (see Section V-B).

The concept of the user interface remains the same as the low-fidelity prototype. There are used screens specified in the low-fidelity prototype but the visual appearance of the user interface is improved to get closer to the final product. The experience is also enhanced by the implementation of simple animations or effects, for example, animation of expanding notification detail or buttons hover effect. There are also changes based on the requirements, comments and suggestions, gained during the low-fidelity prototype evaluation as well as changes reflecting the findings of the evaluation of the low-fidelity prototype. The example high-fidelity screen prototype can be seen in Figure 5.

As described above, there are some changes based on the requirements, comments and suggestions, gained during the low-fidelity prototype evaluation and also changes reflecting findings of the evaluation of the low-fidelity prototype, see section V-A. The main changes are listed below:

***Naming of screen Differential diagnostics sheet***

The naming of the screen *Differential diagnostics sheet* has been changed to *Form of care*, which should eliminate the problems that occurred during low-fidelity prototype evaluation.

***Highlight of notifications***

The notification is highlighted by red color. This change should fix the problem of feeble notification, that was faced by participants during low-fidelity prototype evaluation.

***Unsaved data alert***

The notification about the unsaved data has been added to the prototype. This should eliminate the situation, when users omitted the final confirmation of data change.

***Extended attribute values options***

Text description is added for “other” attribute value option.

Also, an undefined value of attributes, which values are selected through radio buttons is possible.

### Client information print

There has been added a possibility to print out the information about the client.

Home page > List of clients > Jaroslav Novotný

Jan Novák

**Jaroslav Novotný** 12.05.1940 (79 years)

Basic information Biography Activities Medical records Abilities and disabilities Interaction and preferences History

**Basic information**

Name Jaroslav ✓  
 Surname Novotný ✓  
 Birthplace Praha ✓  
 Permanent residence Jungmannova 15, 11000, Praha 1 ✓  
 Birthdate 12.05.1940 ✓  
 Gender - Male - Female ✓  
 Department D ✓  
 Room 111 ✓  
 Key worker Jan Novák ✓  
 Education primary ✓

Contact person

Name & Surname	Relationship	Phone	Email	Action
Ivana Stará	daughter	+420123456789	ivana@stara.cz	✕

Fig. 5. High-fidelity prototype – SCR4a – Client detail – Basic information.

## V. EVALUATION

We evaluated two generations of prototypes of the administration UI with employees of the residential care home. The relatively low number of participants was determined by complicated methods to reach such a specific user group and correspondingly high costs per participant. More information about determining sample size for hard-to-reach user audience can be found in [20], [21].

### A. Administration UI Design Low-Fi prototype

**Participants.** Low-fidelity prototype evaluation was conducted with three participants (P1–P3), all women, average age 33.33 ( $SD = 11.09$ ,  $MIN = 25$ ,  $MAX = 49$ ). All participants work as activation services workers. One of the participants (P2) was previously interviewed in the user research. None of the participants have participated in usability study before. Mean duration of current job title was 1.33 years ( $SD = 0.47$ ).

**Procedure.** The usability study was conducted with employees of the institution for visually impaired people. The test session was under non-laboratory conditions, and it took maximum of an hour. Paper low-fidelity prototype and The Wizard of Oz technique were used. The participants were recorded on a camera in order to log the testing session afterward. Firstly, the participants were informed about the process of the testing session. After the briefing, they filled in a pre-test questionnaire. Then, the system (application) was introduced to them briefly, and they were informed about the purpose of the system. Participants were encouraged to comment aloud their activities. After that, participants were

asked to complete a list of three complex tasks focused on intended typical interaction with the user interface:

- logging in and out
- searching for a client
- viewing information
- editing information
- adding client

After completing all tasks, participants filled in the post-test questionnaire. Finally, participants were asked to share their opinion on the application and on the testing session.

**Results and conclusion.** During the usability study of the low-fidelity prototype, the basic functionality of the user interface was verified.

The usability study also revealed three problems of the proposed design, which must be fixed in the next stages of the design process. The biggest problem faced by all participants was the incomprehensible and misleading term “*Differential diagnostics sheet*”, which was used based on literature analysis. It turned out that none of the participants could imagine anything under this term. This problem can be fixed by renaming the item to “*Form of care*”, which was also suggested by all the participants. Another, quite fundamental problem was feeble notification, which was represented by a tiny black circle. This problem was faced by two participants. The solution to this problem is to make notification more visible. And the last problem, also faced by two participants, was the problem of leaving out the final confirmation of the information editing. The solution is to notify the user, if he/she does not confirm editing and he/she intends to move to another screen. Highlighting the buttons could also help to improve interaction.

The usability study also provided valuable feedback from the participants, and the knowledge gained during testing and post-test interviews can be useful to improve the design of the user interface. Post-test interviews also brought a requirement for new functionality – printing of client information.

The positive information is that all the participants quickly became familiar with the application, even searching for information (attributes) has become easier after a while. All the participants also emphasise the clarity of the proposed design.

### B. Administration UI Design High-Fi prototype

**Participants.** High-fidelity prototype evaluation was conducted with six participants (P1–P6), all women, average age 43.5 ( $SD = 14$ ,  $MIN = 26$ ,  $MAX = 64$ ). Three participants work as activation service workers (P1, P2, P3), two participants as direct care workers (P4, P6) and one participant (P5) works as speech and occupational therapists. Two of the participants (P2, P3) have participated in previous low-fidelity prototype evaluation. Mean duration of current job title was 6.87 years ( $SD = 4.7$ ).

**Procedure.** As same as low-fidelity prototype evaluation, the usability study of the high-fidelity prototype was conducted with employees of the institution for visually impaired people. The test session was under non-laboratory conditions, and it took a maximum of 45 minutes. Participants were using

a computer (MacBook Pro, screen size 13 inches, Google Chrome), and remote computer mouse. The screen of the computer was recorded with sound captured by the internal laptop microphone to log the testing session afterwards. Firstly, the participants were informed about the process of the testing session and after the briefing, they filled in a pre-test questionnaire (the questionnaire contained the same questions as in the low-fidelity prototype evaluation). Then everything went on the same way as evaluation of the low-fidelity prototype, the system (application) was introduced to participants briefly, they were informed about the purpose of the system and they were encouraged to comment aloud their activities. After that, participants were asked to complete a list of three complex tasks focused on intended typical interaction with the user interface. After completing all tasks, participants filled in the post-test questionnaire (the questionnaire contained the same questions as in the low-fidelity prototype evaluation). Finally, participants were asked to share their opinion on the application and the testing session.

**Results and conclusion.** During the usability study of the high-fidelity prototype, the more advanced functionality, compared to low-fidelity evaluation, of the user interface was verified.

The usability study revealed three problems of the proposed design, which may be fixed before the final product will be released. The biggest problem faced by all participants was a difficulty of searching for individual attributes. The problem is caused by a large number of attributes and also the ambiguity of distribution into the individual categories. The problem can be fixed by adding a search box, which can be used for the individual attributes search. Another problem is the unclear difference between screens *Activities* and *Activities of daily living*. Five participants out of six have faced this problem. During the first walkthrough through the application, screens naming may be confusing and also, *Activities of daily living (SCR4b2)* screen is hidden by default as a *Biography* sub-page. The solution is to rename the screen *Activities* to improve user orientation in the application, for example, the screen could be renamed to "*Activities of the client*". And the last problem, which was faced by three participants, were small and feeble radio buttons and checkboxes. The participants complained about the size of the radio buttons and checkboxes and these interaction elements were also very feeble in their point of view. This problem can be fixed by increasing the size of active elements of the interaction.

The knowledge gained during testing and post-test interviews can be very useful to improve the design before releasing the final product. Post-test interviews also brought a requirement for new attribute – information when the client joined the residential care home.

Despite all the problems faced by the participants during the application evaluation, very positive feedback on the application's clarity was obtained.

## VI. DISCUSSION

The evaluation of both generations of the prototypes indicated that the user interface is generally clear for employees of the residential care home. Although some employees in the user study indicated that they like the paper form of the biographical sheet, the electronic form can bring further benefits. It would be possible to track the progress of clients disease by analyzing the biographical sheet and activities performed with the client. Also, the electronic form will bring better privacy security by authorized access and by tracking the access to the biographical sheet.

The validity of the user study and evaluation is limited due to relatively low number of participants and the fact the research has been conducted in a single residential care institution. However, there is a strong evidence that older adults, especially those with cognitive issues like Alzheimer's disease require personalized care. Also, interaction with technologies needs to be adapted to their needs, preferences and changing abilities. This fact is also supported by literature, e.g., [14].

Mutual connection of the personal psychobiography [5], [6] and user profile in one user model could bring various benefits for interaction with technologies. It would be possible to adjust a user interface according to the needs and preferences of a particular client. Properties like rate of speech, information complexity, the volume can be automatically considered. Also, a significant part of the clients has some remaining residual sight that can be used in interactive scenarios. Knowing particular client abilities can improve interaction significantly. For instance, some clients have central vision quite intact, others have peripheral vision only while some can only sense high contrast patterns. Also, the personal psychobiography can be used to make the interaction more natural and personal. Data in the user model can be used for client identification utilizing biometry. Upon this, it is possible to address the client by her/his name in an appropriate form. Also, technological applications like interactive indoor orientation system can provide personalized information like instructions on how to navigate to one's own room.

## VII. CONCLUSION

For the specific user group of older adults with vision impairments, we proposed a solution that consists of the user model structure, corresponding administration user interface and appropriate API. The user model design connects aspects of personal psychobiography that is already successfully applied in the gerontological care and aspects that are important for interaction with technologies.

It is the subject of the future work to fill the user model with data of real clients of the residential care home. Then, the model will be evaluated for the purposes of providing personalized daily care. Furthermore, technological applications will use the model for adaptation of user interfaces and interaction. It would be possible to adapt properties like information complexity, speech rate, or volume. Also, it will be possible to provide more personal communication by addressing the client by her or his name in an appropriate way.

Selin and Rossi [22] propose a method to design safer buildings based on information models. They propose to use *Building Information Modelling (BIM)* for simulation of various situations individuals can deal with in an interior, including evacuation. They use a gaming engine and artificial intelligence to simulate individual with different capabilities. It is the subject of the future work to incorporate modeling of the indoor environment and the ability to plan routes based on needs, preferences, and capabilities of individual users.

#### ACKNOWLEDGMENT

This research has been supported by the TACR research program TE01020415 and the project RCI (reg. no. CZ.02.1.01/0.0/0.0/16\_019/0000765) supported by EU and by the project Navigation of handicapped people funded by grant no. SGS19/178/OHK3/3T/13 (FIS 13139/161/1611937C000).

#### REFERENCES

- [1] R. R. Bourne, S. R. Flaxman, T. Braithwaite, M. V. Cicinelli, A. Das, J. B. Jonas, J. Keeffe, J. H. Kempen, J. Leasher, H. Limburg *et al.*, "Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis," *The Lancet Global Health*, vol. 5, no. 9, pp. e888–e897, 2017. doi: 10.1016/S2214-109X(17)30293-0. [Online]. Available: [https://doi.org/10.1016/S2214-109X\(17\)30293-0](https://doi.org/10.1016/S2214-109X(17)30293-0)
- [2] World Health Organization, *World report on ageing and health*. World Health Organization, 2015.
- [3] Český statistický úřad. (2014) Výběrové šetření zdravotně postižených osob - 2013. [Online]. Available: <https://www.czso.cz/csu/czso/vyberove-setreni-zdravotne-postizenych-osob-2013-qacmwuvwsb>
- [4] M. Macik, I. Maly, J. Balata, and Z. Mikovec, "How can ict help the visually impaired older adults in residential care institutions: The everyday needs survey," in *2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*. IEEE, 2017. doi: 10.1109/CogInfoCom.2017.8268234 pp. 000 157–000 164. [Online]. Available: <http://dx.doi.org/10.1109/CogInfoCom.2017.8268234>
- [5] E. Böhm and P. Sochová, *Psychobiografický model péče podle Böhma*. Mladá fronta, 2015.
- [6] E. Procházková, *Práce s biografii a plány péče*. Mladá fronta, 2014.
- [7] World Health Organization, *International statistical classification of diseases and related health problems*. World Health Organization, 2004, vol. 1.
- [8] World Health Organization and others, *International classification of functioning, disability and health: ICF*. Geneva: World Health Organization, 2001.
- [9] J. O. Wobbrock, K. Z. Gajos, S. K. Kane, and G. C. Vanderheiden, "Ability-based design," *Communications of the ACM*, vol. 61, no. 6, pp. 62–71, 2018. doi: 10.1145/3148051. [Online]. Available: <http://dx.doi.org/10.1145/3148051>
- [10] M. Peißner, D. Janssen, and T. Sellner, "Myui individualization patterns for accessible and adaptive user interfaces," in *The First International Conference on Smart Systems, Devices and Technologies*, 2012, pp. 25–20.
- [11] D. Heckmann, T. Schwartz, B. Brandherm, M. Schmitz, and M. von Wilamowitz-Moellendorff, "Gumo—the general user model ontology," in *International Conference on User Modeling*. Springer, 2005, pp. 428–432.
- [12] D. Heckmann, T. Schwartz, B. Brandherm, and A. Kröner, "Decentralized user modeling with userml and gumo," in *Decentralized, Agent Based and Social Approaches to User Modeling, Workshop DASUM-05 at 9th International Conference on User Modelling, UM2005*, 2005, pp. 61–66.
- [13] P. Kikiras, V. Tsetsos, V. Papataxiarhis, T. Katsikas, and S. Hadjiefthymiades, "User modeling for pedestrian navigation services," in *Advances in ubiquitous user modelling*. Springer, 2009, pp. 111–133.
- [14] J. Hoey, C. Boutilier, P. Poupart, P. Olivier, A. Monk, and A. Mihailidis, "People, sensors, decisions: Customizable and adaptive technologies for assistance in healthcare," *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 2, no. 4, p. 20, 2012. doi: 10.1145/2395123.2395125. [Online]. Available: <http://dx.doi.org/10.1145/2395123.2395125>
- [15] L. Yao, B. Benatallah, X. Wang, N. K. Tran, and Q. Lu, "Context as a service: realizing internet of things-aware processes for the independent living of the elderly," in *International Conference on Service-Oriented Computing*. Springer, 2016. doi: 10.1007/978-3-319-46295-0\_54 pp. 763–779.
- [16] B. Endrstova, M. Macik, and L. Treml, "Reprobooktor: A concept of audiobook player for visually impaired older adults," in *2018 9th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*. IEEE, 2018. doi: 10.1109/CogInfoCom.2018.8639950 pp. 000 063–000 068. [Online]. Available: <https://doi.org/10.1109/CogInfoCom.2018.8639950>
- [17] World Health Organization, "International classification of diseases, 11th revision," Dec 2018, <https://icd.who.int/browse11/l-m/en/#http://id.who.int/icd/entity/1103667651>.
- [18] Balsamiq. Balsamiq wireframes. [Online]. Available: <https://balsamiq.com/wireframes>
- [19] AxureSoftware. Prototypes, specifications, and diagrams in one tool. [Online]. Available: <https://www.axure.com>
- [20] K. Caine, "Local standards for sample size at chi," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2016. doi: 10.1145/2858036.2858498 pp. 981–992. [Online]. Available: <http://dx.doi.org/10.1145/2858036.2858498>
- [21] P. Bacchetti, S. G. Deeks, and J. M. McCune, "Breaking free of sample size dogma to perform innovative translational research," *Science translational medicine*, vol. 3, no. 87, pp. 87ps24–87ps24, 2011. doi: 10.1126/scitranslmed.3001628. [Online]. Available: <http://dx.doi.org/10.1126/scitranslmed.3001628>
- [22] J. Selin and M. Rossi, "The functional design method for buildings (fdm) with gamification of information models and ai help to design safer buildings," in *2018 Federated Conference on Computer Science and Information Systems (FedCSIS)*. IEEE, 2018. doi: 10.15439/2018F162 pp. 907–911. [Online]. Available: <http://dx.doi.org/10.15439/2018F162>

# Gamified Augmented Reality Training for An Assembly Task: A Study About User Engagement

Diep Nguyen  
UniTyLab  
Heilbronn University  
Max-Planck-Strasse 39  
74081 Heilbronn, Germany  
Email: diep.nguyen@hs-heilbronn.de

Gerrit Meixner  
UniTyLab  
Heilbronn University  
Max-Planck-Strasse 39  
74081 Heilbronn, Germany  
Email: gerrit.meixner@hs-heilbronn.de

**Abstract**—Augmented Reality and Gamification are displaying beneficial effects to enhance user experience and performance in many domains. They are widespread across many areas like education, industrial training, marketing, and services. However, the idea of combining the two approaches for an innovative training instrument is fairly new, especially in assembly training. Moreover, learning about the effects of gamification on human, user engagement, in particular, is a complicated subject. There have been several efforts toward this direction, yet the overall situation is still nascent. In this work, we present a gamified augmented reality training for an industrial task and investigate user engagement effect while training with the gamified and the nongamified system. The result shows that people perform better and engage to a greater degree in the gamified design.

## I. INTRODUCTION

**A**UGMENTED Reality (AR) is growing stronger than ever. Market research predicts a 70 to 75 billion revenue for AR by 2023 [5] and by 2019 AR for training, in particular, will take place in 20% of large enterprise businesses [6]. AR is the novel technology which superimposes virtual objects upon the real world subjects or environment while enabling real-time interactions [1]. In recent years, AR has captured the research interests in many areas such as education and training [2], [13], assembly and production operations [3], [4]. As a result, the outcome of teaching and learning, skill acquisition and development as well as user experience have shown outstanding beneficial effects.

Gamification, on the other hand, is the term for adapting the design elements which commonly characterize entertainment games into other settings but gaming. While the academic world is still debating on the consensus of definition and scope, the benefits that gamification brings are undeniable. It is not uncommon to say that games are addictive, yet beyond entertainment purposes, they are believed to better life in many aspects [9]. Gamification's ultimate goal is to simulate the fun elements that enhance the user experience, improve worker productivity or advance student engagement. Since gamification is often mistaken with the meaning of the "serious game," which is any full-fledged game that used for other purposes exceeding pure entertainment, we limit the work in this paper to the most widely accepted definition of gamification [10]:



Fig. 1. The GAR design with gamification elements: points, progress bar and signposting.

*“Gamification is the use of game design elements in non-game contexts.”*

Since both AR and gamification already have their certain contribution into the education field, in the context of training especially, it is surprising that gamified AR systems have not been popular for training in the production environment. Accountable for this probably is the fine line between making work fun and making fun of work [7]. Due to the nature of productional work, the misuse of the gamified systems could take away the user's focus attention and result in damages or even injuries. Therefore, here we attempt to form a gamified AR application for an assembly training task following special design requirements for a production environment. Our focus is on the user engagement aspect because it is an important factor contributes to the effectiveness of training.

## II. RELATED WORK

Although the term “gamification” is relatively new, since around 2003, its applications have already widespread across many industrial as well as scholarly fields. Recently in the Gamification 2020 report, Gartner predicted that gamification in combination with emerging technologies will create a significant impact on several fields including the design of employee performance and customer engagement platform [8]. In this context, there are numerous examples of studies for

either AR training or gamified training, yet there was hardly any work on the combination of those.

A recent survey of Seaborn et al. [14] provides a good overview of gamification from a Human-Computer-Interaction perspective in both theoretical and practical lights. The work showed that gamification is primarily practiced in the domain of education, e-learning especially. In the theoretical foundations, there was a dynamic movement towards carving the boundaries between gamification and other similar concepts. The applied research, meanwhile, painted a positive-leaning but mixed picture about the effectiveness of gamified systems. Despite usual expectation, similar gamified designs under different settings returned clashing result over user experience along with performance. The reason was believed to be highly context-specific requirements. Furthermore, learning about the effects of gamification on the human is a complicated subject. The overall effort toward this direction is still nascent.

While the gamified system was well accepted in business contexts, it is not necessarily the case in production training, left alone Augmented Reality training. K. Lee [13] showed that AR for education and training innovation was leaning towards the “serious game” pole while gamification was left outside of the picture. According to Lee, AR games were particularly interested in by both “educators and corporate venues.” A role-playing game for teaching history [11], for example, proved the benefit of enabling students for problem-solving, increasing collaboration and exploration via the virtual identities.

However, whether we like it or not, production training is different from traditional classroom training. When transforming the operational work into a game, a serious game, there will always be a risk of taking the focus away from the task at hand. This is when gamification comes to play as integrating gamification can provide the fun aspect while still keeping the workers’ full attention on the operative job [12].

Probably the most well-known gamification in production is a series of works from Korn et al. [15], [16], [17], [12]. The center of his works is to evaluate users’ acceptance of gamification in modern production environments. Different designs, “Circles & Bars” and “Pyramid,” were proposed [12]. Both designs were used to visualize work steps as well as their sequences. Color-coded from dark green to yellow, orange and red is employed to indicate user specific time progression. Later on, they were projected into users’ working space as an assistive application for impaired individuals. The result indicated a good acceptance level for gamification designs and the “Pyramid” approach was favorable in general. While the study showed a promising outcome, it focused on user acceptance and did not measure the quantitative factor of gamification on task completion time and error rates.

### III. IMPLEMENTATION

In this section, we present the implementation of the application under study. A process of replacing the battery for a robot arm was implemented based on the instruction manual of the Mitsubishi Industrial Robot RV-2F Series [18]. The application ran on the Microsoft HoloLens [19]. Two

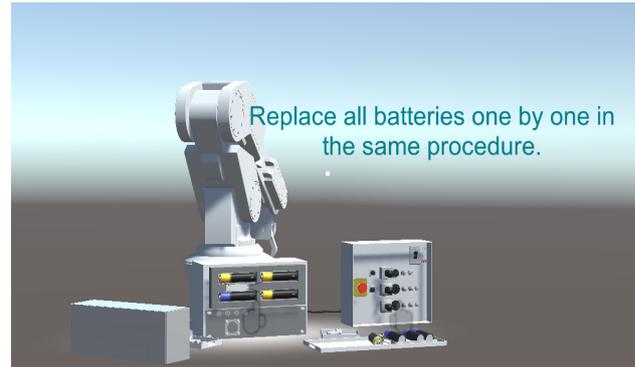


Fig. 2. The NGAR design with no gamification elements. Only text instruction was provided.

prototypes were made, one with the gamification design and the other without. The designs were named Gamification AR (GAR) and Non-Gamification AR (NGAR) according to their characteristics. Due to Microsoft HoloLens small field of view, around 35 degrees, here we provide the user interfaces captured from Unity Editor to showcase the whole scene setup. Figure 1 and Figure 2 illustrate the GAR and NGAR design respectively.

#### A. The application

The process for changing the battery was identically built for both prototypes. There were 21 actions made up 10 steps. Disassembling the cover of the battery compartment, for example, included two steps of removing the screws and removing the cover. While removing each of the screws was counted as an action.

For navigating the process, we augmented the instruction text for each step as a head-up display which was always facing the user at the top right corner of the user view. An instruction manager was used to control the flow of text visualization. The requirement from the instruction manual specified that the steps of the process had to be performed in a fixed order that’s why only one instruction was displayed at a time. The next instruction triggered when the user carried the current step correctly.

Two main interaction types were used to simulate different interactions. Air tap [19] was used for interacting with static objects (e.g. pressing a button) while we utilized drag and drop for assembling actions (e.g. removing the screw). Similar to the real working space, disassembled objects were designed to be placed at a specific location. For instance, the screws needed to be placed inside a designated tray instead of dropped on the floor.

To simulate a sense of reality, sounds such as robot arm were running or turned off were used.

#### B. Gamification Design

The game design elements were implemented only for the GAR version. It allows to isolate and analyze the effect

of gamified system on the user. This could be reflected by comparing the outcome of the two experiments.

As a result of Korn's investigation [12], gamification in the production environment has its own specific requirements. To avoid resistance from users or the potential of taking away their main focuses, we followed the identified requirements in designing gamified application for production settings. First, "keep the visualization of gamification simple." This focuses mainly on avoiding animation, moving elements and using complex graphical structures. The second and third requirements come together as "avoid explicit interaction with gamification elements" and "support implicit interaction with gamification elements." For that matter, in our designs we did not ask for any user's effort to direct input or reach out to the gamified items.

1) *Point System*: The point system was built based on users' actions. There was a maximum of 21 points according to 21 actions. Points were rewarded to the user when the action was done. As the first attempt to study the effect of gamification design on user engagement, we did not implement a complex point system with losing points or rewarding extra points at this stage.

2) *Progress Bar*: While the points were based on actions, progress bar visualized the steps. As stated as one of the requirements, the user interface was intentionally kept simple with only one color. Additional text was in place for indicating the percentage.

3) *Signposting*: Signposting aims to direct the user in the right direction. While users without background knowledge could be confused with the mechanical part names (e.g. Controller box), signposting highlighted the part corresponding to the currently displayed instruction. It provided the "just-in-time" hints for the trainees, especially the totally beginner one.

#### IV. EXPERIMENT DESIGN

The experiment was conducted to investigate how gamification in AR training impacts user engagement and performance. The studies for both conditions (GAR and NGAR) took place in the same room at our research laboratory. To avoid the learning effect, we employed the between-group design in which each participant randomly exposed to only one design, either GAR or NGAR.

Due to the fact that Microsoft HoloLens requires specific hand gestures for interaction, the participants were asked if they have experience with this device. In the case of none, the participant used the default HoloLens "Learn gesture" application. This was especially important because the main task could not be carried on without this step. Before the experiment, regardless of the HoloLens experience, we repeated the main information about the interactive gestures to all participants.

Once the participants were confident interacting with the device, the main experiment task proceeded. When the user hit the "Start" button at the first scene of the application, the timer for measuring task completion time was started until the last step completed.

As we focused on the user engagement we used a post-study questionnaire with the refined User Engagement Scale (UES) [20]. UES is a five-point rating scale: strongly disagree, disagree, neither disagree nor agree, agree and strongly agree, respectively from 1 to 5 point. Given the task was not complicated, the level of fatigue after that was expected not to be high so that we decided to use the UES long form (UES - LF). The UES - LF consists of 30 items covering 4 factors:

- 1) FA: Focused Attention
- 2) PU: Perceived Usability
- 3) AE: Aesthetic Appeal
- 4) RW: Reward Factor

As constructed in the guide to use of UES, all items were randomized and the indicators (e.g. AE.1) were not visible to the users.

#### V. RESULTS

Most of the participants reported having little or none experience with AR technology, in particular, Microsoft HoloLens, before this experiment. So, a potential novelty effect when initially establishing interaction with new technology might influence the research result. The test population was 22 participants with 11 regarding each condition. Participants ages vary from 18 to 34 years old, 15 male and 7 female subjects. Although some unease and uncertainty were expressed at the beginning, all participants were more certain after the learning gesture phase.

Figure 3 displays that the GAR design was rated better in all sub categories. In general, it was clearly preferred to the NGAR approach. The overall Engagement score was 15.2 (SD=1.8) in GAR and 13.3 (SD=3.5) in NGAR. However, this did not make up a statistically significant difference between the two groups. Table I provides the results in more detail, looking at the average score, standard deviation and also the result of a t-test for both the overall engagement score and its factor.

The standard deviation in the overall user engagement score was much lower in the GAR design (SD=1.8), versus SD=3.5 in NGAR, which shows that the GAR subjects more homogeneously perceived the result throughout the group. This tendency, lower standard deviation, remained true for all four subfactors in the GAR design as shown in Figure 3. On the other side, the opinions of NGAR subjects seem to be more diverse.

Looking at the training performance, the difference regarding average task completion time (in seconds) between the two study conditions is statistically significant. The t-test resulted in  $p < 0.032$ . The average time was 306.9 (SD=123.2) and 439.5 (SD=134.4) for GAR and NGAR groups respectively. This positive outcome probably directly influenced by the signposting design element.

#### VI. DISCUSSION AND FUTURE WORK

As a preliminary result, this work demonstrates the potential of gamified AR training for assembly tasks in improving user engagement and performance. Nevertheless, there is a need for

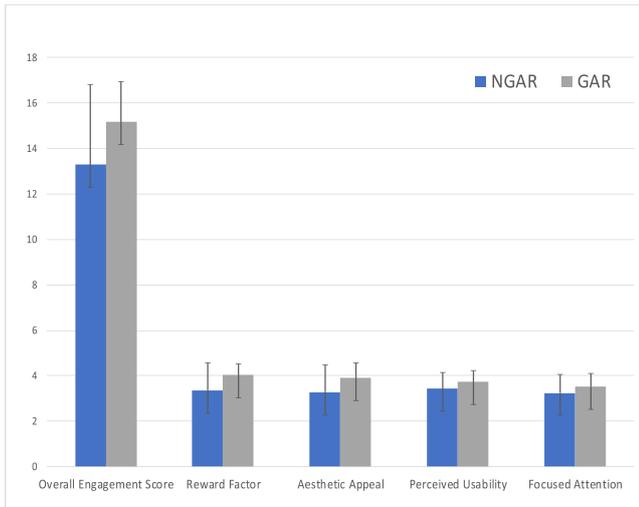


Fig. 3. User Engagement Score as a bar chart with indicated standard deviations.

TABLE I  
COMPARISON OF USER ENGAGEMENT SCORE

Factor	Mean Score (SD)		p value
Design	GAR	NGAR	GAR vs. NGAR
Focused Attention	3.5 (0.6)	3.2 (0.8)	0.418 not significant
Perceived Usability	3.7 (0.5)	3.4 (0.7)	0.281 not significant
Aesthetic Appeal	3.9 (0.7)	3.3 (1.2)	0.162 not significant
Reward Factor	4.0 (0.5)	3.4 (1.2)	0.128 not significant
Overall Score	15.2 (1.8)	13.3 (3.5)	0.153 not significant

further investigation focusing on both short-term and long-term training effectiveness. A consideration over skills and knowledge acquisition should be taken into account. To serve this goal more complex tasks should be implemented with a higher level of gamification, different training levels and challenges design for individual specific demands for example.

As we focused on the improvement of user engagement in gamified AR training, we did not take in to account the isolated effect of how each game design elements affects the user. As mentioned in the Related Work, gamification design is highly context-specific so that the next important step will be a qualitative study on how the users perceive different design elements and their impacts.

## VII. CONCLUSION

The use of gamification in combination with AR for production training is still new and its potential needs further exploration. In this paper, we developed a gamified training for an assembly task in AR setting and studied its effects on user engagement.

The result showed that the users displayed a higher level of engagement as well as better performance with the support of

gamified AR training. The statistical analysis, though, did not indicate a significant difference.

While the implementation of gamification may not yet fully integrate into the training process, this work certainly contributes to the existing knowledge body of gamified AR training for production domain. This research area also needs a greater amount of works to identify its benefits alongside with how to tackle its challenges.

## REFERENCES

- [1] T. R. Azuma, "A Survey of Augmented Reality," *Presence: Teleoperators and Virtual Environments*, vol. 6, no. 4, 1997, pp. 355–385.
- [2] M. Billinghurst, H. Kato and I. Poupyrev, "The Magic Book Moving seamlessly between reality and virtuality," *IEEE Computers, Graphics and Applications*, vol. 21, no. 3, May/June 2001, pp. 2–4.
- [3] S. K. Ong, M. L. Yuan and A. Y. C. Nee, "Augmented reality applications in manufacturing: a survey," *International Journal of Production Research*, vol. 46, 2008, pp. 2707–2742.
- [4] W. Friedrich, "ARVIKA: Augmented Reality for Development, Production and Service," *The 1st International Symposium on Mixed and Augmented Reality (ISMAR)*, 2002, pp. 3–4.
- [5] For AR/VR 2.0 to live, AR/VR 1.0 must die, *Digi-Capital*, <https://www.digi-capital.com/news/2019/01/for-ar-vr-2-0-to-live-ar-vr-1-0-must-die/>. [Retrieved May 2019]
- [6] Transform Business Outcomes With Immersive Technology, *Smarter with Gartner*, <https://www.gartner.com/smarterwithgartner/transform-business-outcomes-with-immersive-technology/>. [Retrieved May 2019]
- [7] S. Dale, "Gamification: Making work fun, or making fun of work?," *Business Information Review*, vol. 31, no. (2), 2014, pp. 82–90.
- [8] Gamification 2020: What Is the Future of Gamification?, <https://www.gartner.com/en/documents/2226015>. [Retrieved May 2019]
- [9] J. McGonigal, "Reality is broken: Why games make us better and how they can change the world," New York: Penguin, 2011.
- [10] S. Deterding, D. Dixon, R. Khaled and L. Nacke, "From Game Design Elements to Gamefulness: Defining Gamification," *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, MindTrek 2011*, vol. 11, 2011, pp. 9–15.
- [11] K. L. Schrier, "Revolutionizing history education : using augmented reality games to teach histories," Institute of Technology. Dept. of Comparative Media Studies, Massachusetts, 2005.
- [12] O. Korn, M. Funk and A. Schmidt, "Design approaches for the gamification of production environments," in *Proceedings of the 8th International Conference on Pervasive Technologies Related to Assistive Environments*, ACM, New York, NY, USA, 2015, pp. 1–7.
- [13] K. Lee, "Augmented Reality in Education and Training," *TechTrends*, vol. 56, 2012, pp. 13–21, <https://doi.org/10.1007/s11528-012-0559-3>
- [14] K. Seaborn and D. I. Fels, "Gamification in theory and action: A survey," *International Journal of Human Computer Studies*, vol. 74, 2015, pp. 14–31.
- [15] K. Oliver, "Industrial Playgrounds. How Gamification Helps to Enrich Work for Elderly or Impaired Persons in Production," in: *Proceedings of the 4th ACM SIGCHI Symposium on Engineering Interactive Computing Systems*, New York, 2012, pp. 313-316.
- [16] O. Korn, M. Funk, S. Abele, A. Schmidt and T. Hörz, "Context-aware Assistive Systems at the Workplace. Analyzing the Effects of Projection and Gamification," in: *PETRA 14 Proceedings of the 7th International Conference on Pervasive Technologies Related to Assistive Environments*, ACM, New York, NY, USA, 2014.
- [17] O. Korn, M. Funk and A. Schmidt, "Towards a Gamification of Industrial Production. A Comparative Study in Sheltered Work Environments," in: *Proceedings of the 7th ACM SIGCHI Symposium on Engineering Interactive Computing Systems*, ACM, New York, NY, USA, 2015.
- [18] Mitsubishi Industrial Robot RV-2F Series: Instruction Manual Robot Arm Setup & Maintenance, [www.geva-roboter.at/files/rv-2f\\_series\\_robot\\_arm\\_setup\\_\\_\\_maintenance.pdf](http://www.geva-roboter.at/files/rv-2f_series_robot_arm_setup___maintenance.pdf). [Retrieved May 2019]
- [19] Introduction to the HoloLens, <https://msdn.microsoft.com/en-us/magazine/mt788624.aspx>. [Retrieved May 2019]
- [20] H. L. O'Brien, P. Cairns and M. Hall, "A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form," *International Journal of Human Computer Studies*, vol. 112, 2018, pp. 28–39.

# 6<sup>th</sup> Doctoral Symposium on Recent Advances in Information Technology

**T**HE aim of this meeting is to provide a platform for exchange of ideas between early-stage researchers, in Computer Science and Information Systems, 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

- 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

## EVENT CHAIRS

- **Kowalski, Piotr**, Systems Research Institute, Polish Academy of Sciences; AGH University of Science and Technology, Poland
- **Łukasik, Szymon**, Systems Research Institute, Polish Academy of Sciences, AGH University of Science and Technology, 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
- **Bronselaer, Antoon**, Department of Telecommunications and Information at Ghent University, Belgium
- **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
- **De Tré, Guy**, Faculty of Engineering and Architecture at Ghent University, Belgium
- **Fonseca, José Manuel**, UNINOVA, Portugal
- **Fournier-Viger, Philippe**, University of Moncton, Canada
- **Gil, David**, University of Alicante, Spain
- **Herrera Viedma, Enrique**, University of Granada, Spain
- **Hu, Bao-Gang**, Institute of Automation, Chinese Academy of Sciences, China
- **Koczy, Laszlo**, Szechenyi Istvan University, Hungary
- **Kokosinski, Zbigniew**, Cracow University of Technology, Poland
- **Krawiec, Krzysztof**, Poznan University of Technology, Poland
- **Kulczycki, Piotr**, Systems Research Institute, Polish Academy of Sciences, Poland
- **Kusy, Maciej**, Rzeszow University of Technology, Poland
- **Lilik, Ferenc**, Szechenyi Istvan University, Hungary
- **Lovassy, Rita**, Obuda University, Hungary
- **Malecki, Piotr**, Institute of Nuclear Physics PAN, Poland
- **Mesiar, Radko**, Slovak University of Technology, Slovakia
- **Mora, André Damas**, UNINOVA, Portugal
- **Noguera i Clofent, Carles**, Institute of Information Theory and Automation (UTIA), Academy of Sciences of the Czech Republic, Czech Republic
- **Pamin, Jerzy**, Institute for Computational Civil Engineering, Cracow University of Technology, Poland
- **Petrik, Milan**, Czech University of Life Sciences Prague, Faculty of Engineering, Department of Mathematics, Czech Republic
- **Ribeiro, Rita A.**, UNINOVA, Portugal

- **Sachenko, Anatoly**, Ternopil State Economic University, Ukraine
- **Samotyy, Volodymyr**, Lviv State University of Life Safety, Ukraine
- **Szafran, Bartlomiej**, Faculty of Physics and Applied Computer Science, AGH University of Science and Technology, Poland
- **Tormasi, Alex**, Szechenyi Istvan University, Hungary
- **Wei, Wei**, School of Computer science and engineering, Xi'an University of Technology, China
- **Wysocki, Marian**, Rzeszow University of Technology, Poland
- **Yang, Yujiu**, Tsinghua University, China
- **Zadrozny, Slawomir**, Systems Research Institute, Poland
- **Zajac, Mieczyslaw**, Cracow University of Technology, Poland

# Sustainable Management of Marine Fish Stocks by Means of Sliding Mode Control

Katharina Benz  
Institute of Product  
and  
Process Innovation  
Leuphana University of Lueneburg  
Volgershall 1,  
D-21339 Lueneburg, Germany.  
Email: katharina.benz@stud.leuphana.de

Claus Rech  
Institute of Product  
and  
Process Innovation  
Leuphana University of Lueneburg  
Volgershall 1,  
D-21339 Lueneburg, Germany.  
Email: clausrech@gmail.de

Paolo Mercorelli  
Institute of Product  
and  
Process Innovation  
Leuphana University of Lueneburg  
Volgershall 1,  
D-21339 Lueneburg, Germany.  
Email: mercorelli@uni.leuphana.de

**Abstract**—This paper deals with a possible approach to controlling marine fish stocks using the prey-predator model described by the Lotka-Volterra equations. The control strategy is conceived using the sliding mode control (SMC) approach which, based on the Lyapunov theorem, offers the possibility to track desired functions, thus guaranteeing the stability of the controlled system. This approach can be used for sustainable management of marine fish stocks: through the developed algorithm, the appropriate number of active fishermen and the suitable period for fishing can be determined. Computer simulations validate the proposed approach.

## I. INTRODUCTION AND MOTIVATION

Marine ecosystems provide humanity with a multitude of goods and services, including water quality, flood control and food supply, all of which are critical for human welfare. Since the human population is growing continuously, the demand for these goods and services is also increasing and progressively exerting more pressure on aquatic ecosystems. As many fish species migrate frequently and the oceans are mostly defined as public areas, the definition of clear boundaries and property rights regarding marine resources is rather complicated. As a result, most natural resources exploited by the fishing industry are defined as common-pool resources. This has resulted in many pelagic ecosystems experiencing high levels of depletion and overexploitation [2], with 46 % of European community fish stocks currently below their minimum biological level (European Environment Agency, [1]). The increasing intensity of human fishing activities in turn diminishes the biodiversity within the affected systems, which is positively correlated with the provision of the goods and services of the ecosystem that are of benefit to the human population, see [5]. Levels of biodiversity have been shown to determine the stability of marine ecosystems and their ability to recover. Consequently, Worm et al. suggest that business as usual in the fishing industry could potentially threaten global food security and water quality, as well as ecosystem resilience, and thus jeopardise present and future generations, see [5]. The observed trend is thus of increasing concern, so the topic of the conservation and restoration of aquatic biodiversity through sustainable fishery management is increasingly

visible in scientific and political agendas. The United Nations has included this issue in its sustainable development goals, dedicating goal number 14 to the conservation and sustainable usage of the planet's oceans, seas and marine resources, [4]. The successful implementation of this goal includes the adaptation of sustainable methods to manage marine and coastal ecosystems in order to avoid significant adverse effects, which is indicated by the proportion of national economic zones following ecosystem-based approaches. By 2020, the United Nations aims to regulate destructive fishing activities and end overfishing, alongside implementing a science-based management approach to restore natural fish stocks (United Nations, 2019). In addition, the European Union has conducted several reforms of the Common Fisheries Policy (CFP), establishing different approaches to attempt to bring the situation under control, with the goal of reaching and maintaining a sustainable level of fish in the oceans and in fishermen's nets. As common practice in this field, scientists estimate the existing level of fish stocks within an area and suggest a number of total allowable catches (TACs) to political fishery ministers. In turn, those ministers try to bargain and receive the highest shares for their regions, which often leads to the amount of TACs exceeding the maximum level recommended by scientists, rather than levels being allocated for mutual benefit and optimal conservation purposes. As a result, the methods of the EU are rather unsuccessful for maintaining a sustainable yield of fish and achieving the targets adopted by all member states of the United Nations: as [3] claims, the decision-making process within the catch allocation should be managed by scientists rather than by politicians. One possible approach to enhancing this decision-making process and expanding it based on an independent and objective component, driven by scientific data, is to translate the observed ecosystem into a mathematical model using MATLAB and simulate them with the integrated tool Simulink. Thus, this paper aims to offer a first attempt at exploring how MATLAB and Simulink can be utilised to facilitate the implementation of sustainable management approaches in the fishing industry through strategic policy testing. The software will be used

to formulate a simple mathematical description of a marine ecosystem based upon the prey-predator system represented in the Lotka-Volterra equations. A number of papers dealing with simulated prey-predator systems have been published previously; however, adaptation of the model to a marine ecosystem including fish stocks and human fishers has not yet been covered. In order to simulate the consequences of various possible policies through different controllers, these have been incorporated into the code to eventually reach and maintain a certain setpoint equal to the maximum sustainable yield of fish. In terms of the proposed control technique, sliding mode control (SMC) is taken as one of the first possible approaches. In fact, the controllers obtained by an SMC approach show robust properties with respect to parameter uncertainties, as well with respect to more general dynamic uncertainties and to unknown signals. Another application for which SMC has suitable qualities is the field of fault-tolerant control (FTC). In this area, due to intrinsic robustness, SMC models are able to overcome faults and uncertainties. The paper is organised in the following way. In Section II the Lotka-Volterra model is presented. Section III is devoted to the control design performed using SMC. Section IV presents the obtained results and the paper ends with the conclusions drawn.

## II. MODEL DESIGN

The designed model is inspired by the ecological concept of the prey and predator relationship. This concept was formulated by Lotka and Volterra, and is based upon different mathematical theorems.

### A. Lotka-Volterra equations

The assumptions of Lotka and Volterra are taken as a basis to describe the relationship between natural fish stocks and the fishing activities of humans. Lotka and Volterra first describe the population dynamics of two species in a prey and predator relationship through two first-order nonlinear differential equations, as follows:

$$\frac{dx(t)}{dt} = \alpha x(t) - \beta x(t)y(t), \quad (1)$$

$$\frac{dy(t)}{dt} = \delta x(t)y(t) - \gamma y(t), \quad (2)$$

where  $x(t)$  represents the number of prey and  $y(t)$  represents the number of predators.  $\frac{dx(t)}{dt}$  and  $\frac{dy(t)}{dt}$  represent the growth rates of the populations based on the respective changes within their population sizes over time, which is denoted by the term  $t$ .  $\alpha, \beta, \delta, \gamma$  are positive real parameters and describe the interaction between the two populations. The expression (1) represents the dynamics of the prey population, which are calculated by subtracting the rate of predation from the population's intrinsic growth rate. Since it is assumed that the prey has an unlimited food supply, its population grows exponentially if the population of predators and the rate of predation equal zero, which is expressed by the term  $\alpha x(t)$ . In turn, the rate of predation upon the prey is assumed to be proportional to  $\beta x(t)y(t)$ . Thus, if either  $x(t)$  or  $y(t)$  equals

zero, there is no predation.

Equation (2) describes the dynamics of the predator population, which are determined by the rate at which it consumes the prey population, minus its intrinsic death rate. Since the growth rate of the predator population does not necessarily equal the rate of predation of the prey, it is expressed by  $\delta x(t)y(t)$ , which is similar but not equal to the term representing the rate of predation in Eq. (1). In this equation,  $\gamma y(t)$  denotes the loss rate of the predator population due to natural death or emigration. This results in an exponential decay if there is no prey available to be consumed. Since the main objective of designing this new approach is to achieve and maintain sustainable levels of fish stocks and harvests alike, an equilibrium point between the two populations is intended. This point is reached if:

$$\frac{dx(t)}{dt} = 0, \quad (3)$$

$$\frac{dy(t)}{dt} = 0. \quad (4)$$

As a result, putting the corresponding equations also equal zero, wherefore one has:

$$0 = \alpha x(t) - \beta x(t)y(t), \quad (5)$$

$$0 = \delta x(t)y(t) - \gamma y(t). \quad (6)$$

These equations yield two different solutions. One solution states that both populations become extinct:

$$x(t) = 0, \quad y(t) = 0. \quad (7)$$

Given the second solution, a fixed point can be achieved at which both populations sustain their current non-zero numbers, depending on the settings of the four parameters  $\alpha, \beta, \delta, \gamma$ . This yields:

$$y(t) = \frac{\alpha}{\beta}, \quad (8)$$

$$x(t) = \frac{\gamma}{\delta}. \quad (9)$$

## III. SLIDING MODE CONTROL

As the goal of the simulation is to realise and establish sustainable fishing activities in order to ensure the continuity of both marine ecosystems and the human species, the current situation of overfishing and ocean depletion has to be stopped and managed in a way that enables fish stocks to recover. Therefore, the error between the desired setpoint, being the equilibrium point of the fishery system, and the actual value, represented by the current level of fish, has to be ascertained, harmonised and stabilised. This is explored through application of the Lyapunov theorem. With zero being the intended value for  $\dot{x}(t) = f(x, u, t)$ , the theorem defines that if:

$$V(x(t)) > 0, \forall x(t), \quad (10)$$

$$V(0) = 0, \quad (11)$$

the function is positive and if:

$$\dot{V}(x(t)) < 0, \forall x(t) \quad (12)$$

and one has:

$$\dot{x}(t) = f(x, u, t), \quad (13)$$

then  $x(t) = 0$  is an asymptomatic stable point for function  $\dot{x}(t) = f(x, u, t)$ .

In order to reduce the error and harmonise the actual value of fish with the desired value of fish associated with a sustainable population size, an SMC is used as follows:

$$S(t) = (x_d(t) - x(t)) + k_s \int_0^t (x_d(z) - x(z)) dz, \quad (14)$$

where  $k_s$  is a parameter to be designed. Since the V-function is a positive-definite function of  $x(t)$ , it can be employed in the function above. Therefore, one gets:

$$V(S(t)) = \frac{1}{2} S^2(t). \quad (15)$$

Thereupon, the function is differentiated, which yields:

$$\dot{V}(S(t)) = \frac{1}{2} 2S(t)\dot{S}(t), \quad (16)$$

$$= S(t) [\dot{x}_d(t) - \dot{x}(t) + k_s(x_d(t) - x(t))], \quad (17)$$

$$= S(t) [\dot{x}_d(t) - (\alpha x(t) - \beta x(t)y(t)) + k_s(x_d(t) - x(t))], \quad (18)$$

if:  $y(t) = y_{eq}(t) =$

$$\frac{-\dot{x}_d(t) + \alpha x(t) - k_s(x_d(t) - x(t))}{\beta x(t)}, \quad (19)$$

then  $\dot{V}(S(t)) = 0$  and if:

$$y(t) = y_{eq}(t) - \frac{\eta \operatorname{sgn}(S(t))}{\beta x(t)}, \quad (20)$$

with

$$\operatorname{sgn}(S(t)) = \begin{cases} 1 & \text{if } S(t) > 0 \\ 0 & \text{if } S(t) = 0 \\ -1 & \text{if } S(t) < 0, \end{cases} \quad (21)$$

then, if  $\eta > 0$ :

$$\begin{aligned} \dot{V}(S(t)) &= S(t)[- \eta \operatorname{sgn}(S(t))] \\ &= -\eta S(t) \operatorname{sgn}(S(t)) = -\eta |S(t)| < 0. \end{aligned} \quad (22)$$

In order to accelerate the process and reach the desired value more quickly, term  $\lambda S(t)$ , with  $\lambda > 0$ , can be included in the equation. The resulting control law is as follows:

$$y(t) = y_{eq}(t) - \frac{\eta \operatorname{sgn}(S(t))}{\beta x(t)} - \frac{\lambda S(t)}{\beta x(t)}. \quad (23)$$

#### A. Euler Method

Since the system in question has a relatively slow dynamics, it is not intended to measure its state second-by-second, but rather on a monthly basis. Therefore, the equation is discretised according to the Forward Euler method, where  $k$  represents the known counting integer variable, which yields:

$$\begin{aligned} \dot{x}(t) &= \frac{x(k) - x(k-1)}{T_s} \\ &= \alpha x(k-1) - \beta x(k-1)y(k-1) \end{aligned} \quad (24)$$

$$\rightarrow x(k) = x(k-1) + T_s(\alpha x(k-1) - \beta x(k-1)y(k-1)). \quad (25)$$

At this point, the respective equations are integrated into Matlab. With the number of predators and respectively the number of fishermen represented by  $y(t)$ , being the leverage point to control the level of fish stocks in the regarded aquatic ecosystem, Eq. (23) represents one of the main equations in the SMC. Since the goal of the applied controller is to harmonise the desired and actual amounts of fish, measured in kilogram biomass, the desired amount of fish (denoted by  $x_d(t)$ ) and the actual amount of fish (represented by  $x(t)$ ) are the two main data inputs for the equation. Eq. (23) represents the main equation within the SMC strategy.

#### IV. SIMULATION RESULTS

In order to test the designed model it is assumed that a sustainable level of fish stocks is reached at a minimum of 10,000kg of fish. The goal is then to test how the attendance of fishermen affects the dynamics of the prey population and how a meaningful policy designed to regulate the activities of the fishermen could be framed. Figure 1 shows the number of fishermen in a system that is not restricted by political regulations. The line graph shows the development of the number of fishermen over a period of 60 months. In the absence of political regulations, the number of fishermen immediately increases to 1,000 and remains stable over the entire period of time. The line graph depicted in Fig. 2

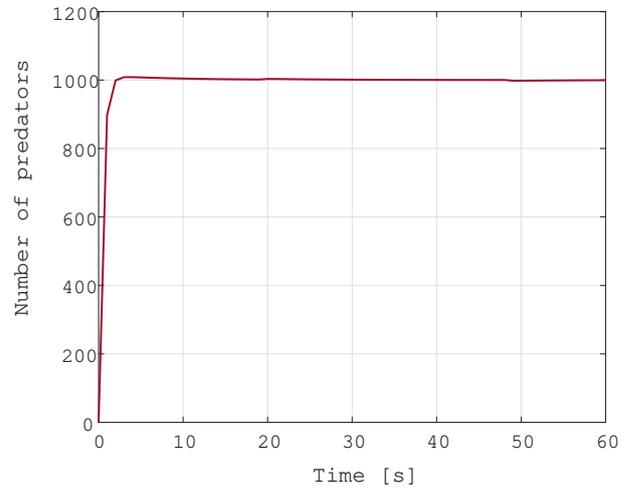


Fig. 1. Number of predators without regulation

shows the corresponding dynamics of the fish population over a period of 60 months, given the same situation that no political regulation of fishing activities exists. In this scenario the amount of fish peaks at 11,000kg after approximately three months and stabilises at the desired amount of 10,000kg after 60 months. In order to test how a political regulation regarding the number of active fishermen affects the system, a hypothetical regulation has been assumed demanding that all

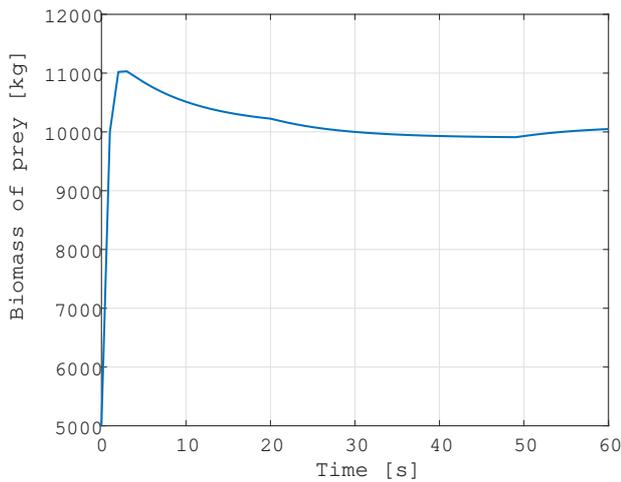


Fig. 2. Biomass of prey without regulation

fishing activities are prohibited between the 5<sup>th</sup> and the 8<sup>th</sup> month of the period in question. This regulation is realised through an if-clause in the m-file of Matlab, as follows:  $if((T < 5)|(T > 8))$

$$y(t) = y_{eq}(t) - \frac{\eta \text{sgn}(S(t))}{\beta x(t)}. \quad (26)$$

As a result, the number of fishermen depicted in Fig. 3 rises to 1,000 and remains at that level until it drops to 0 at the five-month mark. It then remains at 0 until the 8<sup>th</sup> month and temporarily increases to 1,100 after this point. Subsequently, the number slowly decreases again until it returns to a level of 1,000 after 60 months. The consequences of the regulation

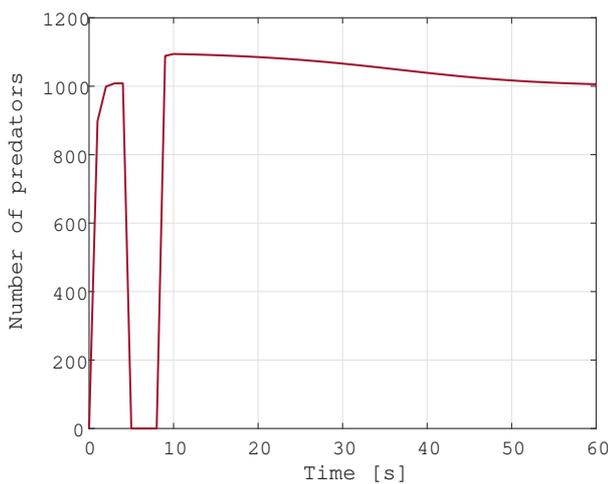


Fig. 3. Number of predators with regulation

regarding the level of fish stocks in kilogram biomass is depicted in Fig. 4. At the beginning of the time period in question, when the number of fishermen is high, the fish

biomass level is at 10,000kg. As soon as the regulation takes effect, the fish biomass increases exponentially, peaking at 17,500kg at eight months. Since the fishermen resume their activities from the 8<sup>th</sup> month onwards, the biomass level decreases again, stabilising at the desired level of 10,000kg after 60 months. The results show that the designed model is

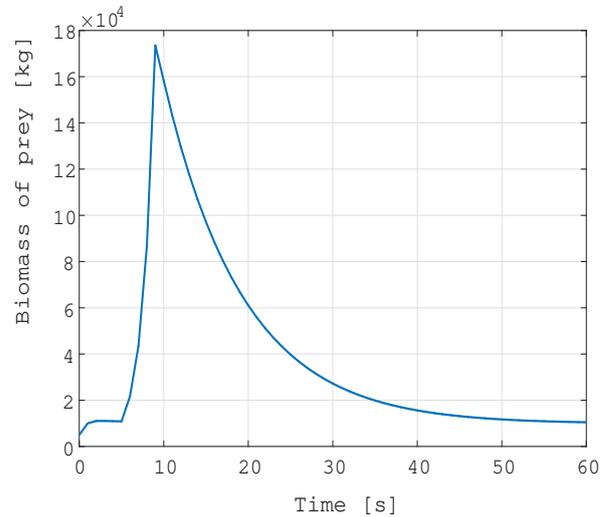


Fig. 4. Biomass of prey with regulation

indeed sensitive to regulatory changes, and that it is able to depict the dynamics of the interdependent populations.

## V. CONCLUSION AND FUTURE RESEARCH

Since the implementation of a regulating if-clause in the m-file yields a reasonable result, the model seems to work and to be appropriate for policy testing in the fishing industry. However, further research will be necessary in order to construct more complex models, and thus more realistic ones, by including additional variables that may influence the system. In addition, appropriate measurements must be taken and the values within the models must be adapted accordingly in order to obtain realistic and meaningful results.

## ACKNOWLEDGMENT

This work was realised within the lectures for the Complementary Studies course at Leuphana University of Lüneburg during the winter semester 2018-2019.

## REFERENCES

- [1] European Environment Agency (EEA) (2010). EU 2010 biodiversity baseline.
- [2] Gordon, H. S. The economic theory of a common-property resource: the fishery. *The Journal of Political Economy*, Vol. 62, No. 2 (Apr., 1954), pp. 124–142
- [3] Leary, B. C., Smart, J. C., Neale, F. C., Hawkins, J. P., Newman, S., Milman, A. C., & Roberts, C. M. (2011) Fisheries mismanagement. *Marine Pollution Bulletin*, 62 (12), pp. 2642–2648.
- [4] United Nations (2019). Sustainable Development Goal 14. [https://sustainabledevelopment.un.org/sdg\\_14](https://sustainabledevelopment.un.org/sdg_14) (20.03.2019).
- [5] Worm, B., Barbier, E. B., Beaumont, N., Duffy, J. E., Folke, C., Halpern, B. S., ... & Sala, E. (2006). Impacts of biodiversity loss on ocean ecosystem services. *Science*, 314 (5800), pp. 787-790.

## An effective industrial control approach

Michal Kostoláni  
Faculty of Electrical Engineering  
and Information Technology  
Slovak University of Technology  
in Bratislava  
Bratislava, Slovak Republic  
Email: michal.kostolani@stuba.sk

Justín Murín  
Faculty of Electrical Engineering  
and Information Technology  
Slovak University of Technology  
in Bratislava  
Bratislava, Slovak Republic  
Email: justin.murin@stuba.sk

Štefan Kozák  
Faculty of Informatics  
Pan-European University  
Bratislava, Slovak Republic  
Email:  
Stefan.kozak@paneurouni.com

**Abstract**—Requirements of increased productivity and flexibility in manufacturing processes reflect the concept Industry 4.0. Essential to achieving these targets is the implementation of intelligent and robust distributed control systems focused on interoperability and scalability. New approaches and technologies based on the Industrial Internet of Things (IIoT), cloud computing and Big Data are an emerging field of industrial control that comprises Internet-enabled cyber-physical devices with the ability to link new smart technologies. With this perspective, manufacturing devices can be easily monitored, operated and controlled even from remote locations. Conventional industrial control approach, that use programmable logic controller (PLC), is enhanced with intelligent industrial gateway IoT 2040 as a hardware and Node-RED software environment in order to provide interoperability, robust and reliable control system. Collected process data and parameters from embedded sensors and other connected devices can be quickly collected, processed, transformed and used for device control from remote production environment. This paper deals with an industrial gateway framework adopting the idea of Internet of Things for the development of robust industrial control approach. The concept is tested in real industrial environment.

### INTRODUCTION

INDUSTRY 4.0 represents the fourth industrial revolution in manufacturing industry with complex automation, cyber-physical systems, data exchange and Industrial Internet of Things (IIoT) principles. The ability of integration and cooperation of intelligent machines, methods and human beings to interact is essential condition for increased productivity and flexibility in manufacturing processes [1].

Interconnection of machines, embedded sensors, digital devices and people continues to extend. The aims of this approach are improved industrial manufacturing processes, efficient ways to operate production plants, services and supervision for industrial installations, reduced operational cost in relation to requirements of improved human safety. The IIoT offers interconnection and intelligence to industrial systems and machines through sensing devices and actuators with ubiquitous networking and computer abilities [1], [2].

The expectation toward industrial applications related to intelligent hardware, software and serviceability are high.

Nowadays, many IoT academic research studies, publications and applications related to the IIoT principles are being developed. The development of intelligent control prototypes has increased mainly due to educational development kits, such as Raspberry Pi, Arduino, etc., that are usually not certified for the use in industry.

The difference between industrial hardware and regular development kit are in UL/ CE certification that indicates conformity with health, safety, and environmental protection standards for industrial continuous operation applications. Secure installation within machine and electrical panels allows minimal industrial IP 20 housing and easy connection with other DIN rail industrial devices, such as relays, power supplies or PLC. For continuous operation in industrial environment must be this hardware built with industrial grade components resistant to vibrations, dust, high temperature and electromagnetic interferences. Regarding communication with other industrial hardware, it must be possible by using secure industrial protocols, such as Profinet, MQTT or ModBus [2].

Developing an effective industrial control approach and IIoT application according to the IIoT principles requires meeting the following requirements on increased [3]:

- Robustness.
- Intelligence.
- Reliability.
- Standardization.
- Safety and security.
- Cost reduction.

The architecture of conventional industrial control system consists of distributed embedded devices, such as PLC that control physical processes and supervisory computer that gather data and command PLC. Modern industrial control systems are distributed, offer higher performance and often connected to the internet or intranet [1], [4].

Machines from different manufactures and on different technological levels often do not use the same communication protocols or programming language. In order to satisfy the need for intelligent and robust control solution suitable for industrial production that harmonizes communication between the various data sources, the

architecture of industrial control system must be enhanced with an industrial certified gateway [4].

IIoT gateway builds the bridge between sensors and actuators of manufacturing device on one hand, and the internet or intranet on the other hand. IIoT gateway moreover comprises other capabilities such as filtering the amount of data or security implementation as depicted in Fig.1.

Industrial gateway represents the reliable open platform for collecting, processing and transferring data in the production environment. It is intelligent gateway between a company's IT department and industrial solutions and it is important to note that the role as an interface can be used in both directions.

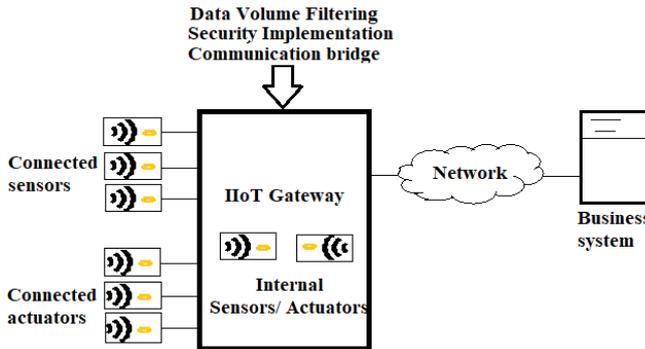


Fig. 1 Industrial IIoT gateway function principle

The paper consists of four parts. In the first part requirements for effective industrial control according to the IIoT methodology are defined. The second part deals with the problem formulation of industrial control systems current state and intelligent gateway. The third part of the paper provides a description of the proposed approach and system implementation in terms of hardware (IoT 2040) and software components (Node RED). The fourth part deals with case study – verification of proposed approach.

#### PROBLEM FORMULATION

This paper proposes a robust, effective and modern approach of process control and remote monitoring system of a real physical model represented by an automated working system in compliance with the main IIoT requirements, interoperability and scalability [4]. This stand-alone system is depicted in Fig.2 and represents an autonomous working station of production line with sorting mechanism and consists of industry certified hardware, such as power supply, engine with frequency inverter, conveyor belt, proximity sensors, pneumatic actuators and programmable control unit. The main functionality of this system is to move packages along the conveyor belt to the sorting mechanism, which divides packages according to the shapes into individual containers. Control of actuators and monitoring of process parameters in manual or automatic operation is only possible by touch panel or control buttons.

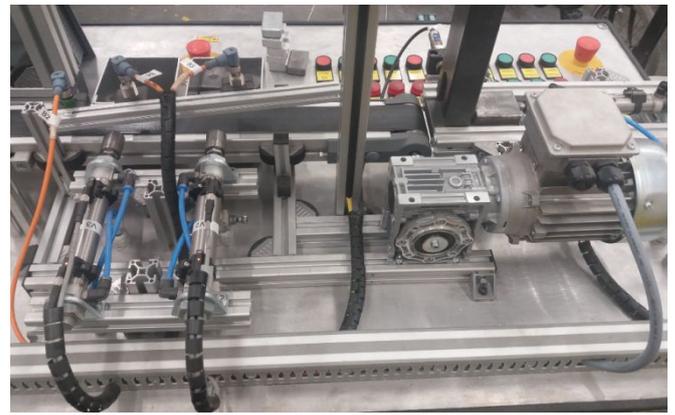


Fig.2 Autonomous stand-alone working station of production line

If there is a system error or breakdown, the operator has to check all process conditions, quit error messages and start the automatic operation through the touch panel.

If there is one system, it is a simple task. If the system is more complex or without the operator's permanent control, there is no possibility to control and improve production processes of the system.

As it is a stand-alone system without access to any network, intranet or internet, interoperability with cloud based systems, other cyber-physical systems or remote monitoring system is not possible. To improve interconnection, scalability, time and cost savings, it is necessary to enhance the stand-alone system with external hardware which enables the connection of working station with a local network [4].

As the best candidate for solving this problem in industrial conditions, Simatic IoT 2040 shown in Fig.3 has been chosen as the open industrial gateway [6]. IoT 2040 is an open platform for collecting, processing and transferring the data between production and IT systems or clouds in the production environment. It is designed for 24/7 operation and the role as an intelligent gateway interface can be used in both directions transferring data. This gateway supports programming languages such as Java, Python or C++, and multiple communications protocols such S7 Protocol, OPC UA, Profinet, TCP IP, MQTT via various interfaces, including RS232/422/485, serial USB, Ethernet or Wi-Fi.



Fig. 3 Industrial gateway IoT 2040

SYSTEM IMPLEMENTATION

Typical industrial control systems consist of distributed embedded devices such as PLC S7-1500 that control physical processes. For effective and robust industrial control approach according to the principles of IIoT, it is essential the hardware and software implementation and configuration of intelligent gateway as depicted in Fig. 4.

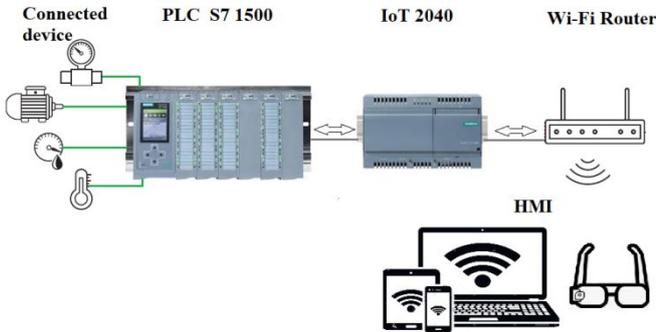


Fig. 4 Implementation of gateway IoT 2040 in to typical control system

In following parts, the exact approach of hardware and software implementation and configuration of IoT 2040 will be analyzed.

A. Setting up S7-1500

Based on environment and technical requirement, PLC S7-1500 was used for real-time control of the stand-alone system described above. Devices, such as sensors, actuators or frequency inverter are connected via Profinet cable directly to the I/O modules of PLC. Communication between these devices is possible via libraries in Totally Integrated Automation (TIA) software [4].

B. Hardware implementation

To the existing real physical model represented by automated working system controlled via PLC the intelligent gateway can be attached horizontally or vertically on a DIN rail or to a wall. Connection to the power supply can only be supported by 9-36 V. The Ethernet cable connected to the port X1 P1 LAN provides the gateway connection with Wi-Fi router. Communication between IoT 2040 and PLC S7-1500 was ensured via Ethernet cable connected to the gateway port X2 P1 LAN as depicted in Fig. 5.



Fig. 5 Hardware implementation of industrial gateway IoT 2040

C. Software implementation of PLC S7 1500

The first step in software implementation of PLC S7 1500 and connected devices, such as sensors and actuators, is to set communication between PLC and I/O modules via device configuration.

Once the communication is established the variables type, address and tag table need to be defined. Networks with appropriate variables from a list are an essential condition for creating automated control system as shown in Fig. 6.

To allow communication with other devices connected to the network, the permission access with PUT/GET communication from remote partner, such as PLC, HMI or OPC must be enabled [7].



Fig.6 Setting of control program with the use of PLC S7 1500

D. Software implementation of IoT 2040 in Node-RED

Industrial gateway IoT 2040 is certified hardware for industrial application. According to the type of application, the image must be created and uploaded to the gateway [5]. The first step for working with industrial gateway IoT 2040 is to set up a micro-SD card with appropriate image and install it to the gateway.

To create IIoT application, the interconnection of the physical layer with Node-RED needs to be established. To start Node-RED, a special command must be typed into the gateway terminal. Once the Node-RED has started, the user interface can be open in the web interface after typing the correct IP address 192.168.43.200:1880 and logging in with user credentials. Basic settings such as IP address of PLC, port, rack, slot, cycle time must be defined to establish the interconnection between correct hardware components. In order to allow remote control and remote visualization of connected devices using gateway IoT 2040, in PLC defined variables need to be defined as well in Node-RED variable list. After completing of basic communication and variables settings, it is possible to continue with the creation of application flows.

The process of flow creation is similar to the PLC flow-based programming. To the input nodes that are located on the left side of Node-RED web editor it is appropriate to assign them in the workspace with a proper variable and other relevant settings [5]. The same logic is used to create output debug nodes. To establish the connection between input and output nodes the wire in web editor is used.

## CASE STUDY

The aim of this case study is to verify the proposed methodology on a diminished form of gearbox production line with sorting mechanism by which parts are divided into individual containers. Individual parts are stored in a magazine and are in regular intervals positioned on the moving conveyor belt. According to the evaluation of the sorting mechanism, individual components are then moved by means of pneumatic cylinder. A HMI (Human Machine Interface) is used as a touch panel. B1, B2, B3 and B4 represent sensors, V1, V2 and V3 represents pneumatic actuators. According to the above mentioned methodology of IIoT application, the data from PLC are used for remote control and visualization. To create a simple application, input nodes connected to the sensor variable (I0.4) and engine variable (Q1.2) are wired with different types of output nodes, for example pneumatic actuators (Q0.0), conveyor belt (M50.1) as depicted in Fig. 7.

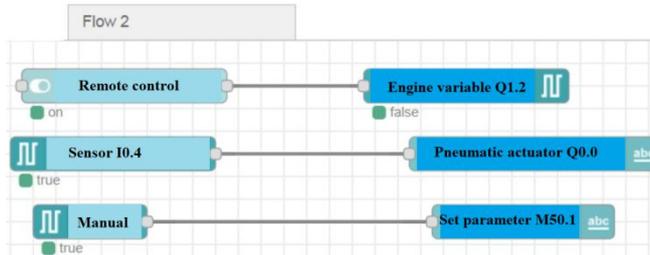


Fig. 7 Settings of input and output nodes in IIoT application in Node-RED

According to the specific requirements of different applications, output information from PLC can be presented in form such as gauge, switch, email, twitter, etc. After all nodes are created, the whole flow needs to be deployed.

Once the whole flow is deployed, remote visualization of the station and its connected components in real time is possible via any smart devices connected to the same local network after typing the correct IP address and log in using safety log in credentials. In the start page of the application, there is the option to set the device to automatic or manual operating mode, as well as to allow the remote control.

In the manual operating mode, using smart phone or other device connected into the same network, is it possible to see remote visualization of variables as depicted in Fig. 8.



Fig. 8 Manual operating mode of mechatronic system via smart phone

To change variables defined in variable list and control the process by changing parameters on switch, gauge or conveyor belt in real time is possible even via smart phone.

## CONCLUSION

In this paper, we managed to interconnect an industrial PLC via intelligent gateway IoT 2040 with web based platform Node-RED. Relevant data for remote control and visualization, such as current state of connected devices or process data can be easily obtained directly from PLC to intranet or internet without the need of additional sensor installation. A case study according to the proposed methodology was applied to verify this approach.

Using this approach, process data is easily accessible even from remote locations, can be analyzed and evaluated to improve production or maintenance processes. This scalable result is a solution that can be extended to the cloud based systems.

Further research and experiments should focus on applying of proposed scalable approach to the more complex industrial equipment for further possibilities of collecting, evaluating obtained data, as well as the use of secured cloud based systems in real time.

## ACKNOWLEDGMENT

This article is supported by the Cultural and Educational Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic, grant No. KEGA 030STU-4/2018 and Scientific Grant Agency of Ministry of Education, Science, Research and Sport of Slovak Republic, grant No. VEGA: 1/0102/18.

## REFERENCES

- [1] M. Wollschlaeger, T. Sauter, a J. Jasperneite, "The Future of Industrial Communication: Automation Networks in the Era of the Internet of Things and Industry 4.0", *IEEE Industrial Electronics Magazine*, roč. 11, č. 1, p. 17–27, mar. 2017. DOI: 10.1109/MIE.2017.2649104
- [2] I. Zolotová, M. Bundzel, and T. Lojka, "Industry IoT Gateway for Cloud Connectivity", in *Advances in Production Management Systems: Innovative Production Management Towards Sustainable Growth*, 2015, s. 59–66. DOI: 10.1007/978-3-319-22759-7\_7
- [3] A. Astarloa, U. Bidarte, J. Jiménez, A. Zuloaga, and J. Lázaro, "Intelligent gateway for Industry 4.0-compliant production", in *IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society*, 2016, p. 4902–4907. DOI: 10.1109/IECON.2016.7793890
- [4] A. Gavlas, J. Zwierzyna, a J. Koziorek, "Possibilities of transfer process data from PLC to Cloud platforms based on IoT", *IFAC-PapersOnLine*, roč. 51, č. 6, s. 156–161, jan. 2018. DOI: 10.1016/j.ifacol.2018.07.146
- [5] J. Skovranek, M. Pies, a R. Hajovsky, "Use of the IQRf and Node-RED technology for control and visualization in an IQMESH network", *IFAC-PapersOnLine*, roč. 51, č. 6, s. 295–300, jan. 2018. DOI: 10.1016/j.ifacol.2018.07.169
- [6] S. Toc a A. Korodi, "Modbus-OPC UA Wrapper Using Node-RED and IoT-2040 with Application in the Water Industry", in *2018 IEEE 16th International Symposium on Intelligent Systems and Informatics (SISY)*, 2018, s. 000099–000104. DOI: 10.1109/SISY.2018.8524749
- [7] M. Kostolani, J. Murin, S. Kozak, "Intelligent predictive maintenance control using augmented reality" in *22<sup>nd</sup> International Conference on Process Control*, 2019, p. 131 – 135.

# PN2ARDUINO - A New Petri Net Software Tool For Control Of Discrete-event And Hybrid Systems Using Arduino Microcontrollers

Erik Kučera, Oto Haffner and Roman Leskovský  
Faculty of Electrical Engineering and Information Technology  
Slovak University of Technology in Bratislava  
Bratislava, Slovakia  
Email: erik.kucera@stuba.sk

**Abstract**—The main aim of proposed paper is the design of new software system for modelling and control of discrete-event and hybrid systems using Arduino and similar microcontrollers. In this article we propose a new tool. This new tool is based on Petri nets and it is called PN2ARDUINO. It offers a capability of communication with the microcontroller. Communication with the microcontroller is based on modified Firmata protocol so control algorithm can be implemented on all microcontrollers that support this type of protocol. The developed software tool was successfully verified for control of laboratory systems. It can also be used for education and also for research purposes as it offers a graphical way for designing control algorithm for hybrid and mainly discrete-event systems. Proposed tool can enrich education and practice in the field of cyber-physical systems.

## I. INTRODUCTION

Development of various systems is a complex discipline that includes many activities, e.g. system design, a specification of required properties, implementation, testing and further development of the system. As these operations are challenging and important for the final product, it is appropriate and necessary to create a model of the system. Development of control methods of discrete-event and hybrid systems belongs to the modern trends in automation and mechatronics. Hybrid system is a combination of continuous and discrete event systems. Control of such systems brings new challenges because it is necessary to join control methods of discrete event systems (where formalism of Petri nets can be helpful) and classic control methods of continuous systems. With good methodology and software module, these approaches can be synergistically combined. This will give us an appropriate and unique control system that allows harmonizing discrete event control methods with the methods of control of continuous systems (e.g. PID algorithms). Effective cooperation of these approaches allows to control hybrid system. This method would be useful in systems where it is necessary to use different control algorithms (for example PID controllers with different parameters) according to the state of the system. The concept of Petri nets is capable of covering a management of these control rules in a very efficient, robust and well-arranged (graphical) way. This paper is aimed to present new

Petri Net tools for modelling and control of discrete-event and hybrid systems. Case studies for control of laboratory fire alarm system and DC motor are also presented.

In [1] author developed an interesting software tool that supports hybrid Petri nets named Visual Object Net++. There a lot of papers (mainly from Romanian author [2]) that describes capabilities of Visual Object Net++. This tool is not open-source and it is not further developed.

As an interesting way of research, a Modelica language and open-source tool OpenModelica appeared. There is a library that supports modelling by Petri nets in this tool. One of the advantages of OpenModelica is that PN model can be connected with other components of Modelica. The first Petri net toolbox was introduced in [3]. An extension of this toolbox was described in [4]. The greater addition to the toolbox was made by the German author who enriches it by a support of extended hybrid Petri nets for modelling of processes in biological organisms [5] and [6]. This tool was developed primarily for commercial tool Dymola and not for OpenModelica, so applicability in scientific research and extensibility is limited. During 2015 the team that developed PNlib published modified version of PNlib that partially worked in OpenModelica. Unfortunately, it was not possible to use OpenModelica for control purposes using microcontrollers because of lack of COM port communication support.

According to the survey made during the described research project, it was realized that it is necessary to develop the own solution for control of discrete and hybrid systems using Petri nets based on microcontrollers as there is a lack of tools that support control of real systems using Petri net formalism.

## II. DESCRIPTION OF DEVELOPED SW TOOL PN2ARDUINO

As it was realized that there is no complex SW solution to support control of discrete event and hybrid system by microcontrollers using High-level Petri nets, it was necessary to develop it. As a basis for such software, PNEditor [7] was chosen. This tool is open-source. The developed extension of this tool is named PN2ARDUINO and is fully tested in [8] and [9]. The main topic of this paper is an introduction to this



Fig. 1. Simple scheme of proposed solution - Petri net's logic in microcontroller



Fig. 2. Simple scheme of proposed solution - Petri net's logic in PC

developed software that can be used for control of discrete event and hybrid systems and its verification on laboratory discrete-event and hybrid system.

There are more concepts of control using Petri nets. Petri net as a control logic is necessary to connect with the controlled system (e.g. using microcontroller). One of the main aspects of the control system design is the question whether the Petri net's logic should be stored in the microcontroller or into the PC (which can communicate with microcontroller). Both approaches have their advantages and disadvantages.

If the Petri net's logic is stored in the microcontroller, the main advantage is the independence of control unit from the software application (program on PC). The Petri net logic is modelled using PC, and then the Petri net is translated into program code which is loaded into the microcontroller. Then PC and microcontroller can be disconnected. The advantage is also the capability of control in real time. Disadvantages are limited computational and memory resources of the microcontroller. Following disadvantage is the need of repeating compiling and uploading the program into the microcontroller (mainly during development phase). The proposed solution is shown in Fig. 1.

When the Petri net's control logic is stored in specialized SW application on PC, this solution gives an opportunity to control the system directly from it. In the microcontroller, only the program with communication protocol is stored. This communication protocol (in our case it is Firmata [10]) is used for communication between PC and microcontroller. This solution eliminates the necessity of recompiling and reuploading the program during development. The next advantage is the elimination of restrictions on computing and storage resources because PC has (in comparison with microcontroller) almost unlimited resources. One of the disadvantages is that the control system cannot react in real time. The proposed solution is shown in Fig. 2.

In Table I, these differences are specified.

New software module PN2ARDUINO was based on the second approach. The Petri net runs on the personal computer. For communication between SW application and microcontroller, the protocol Firmata [10] was used. Firmata is a protocol that is designed for communication between microcontroller and computer (or mobile device like a smartphone, tablet, etc.). This protocol can be implemented in firmware of various



Fig. 3. PN2ARDUINO - Use-case diagram

microcontrollers. Mostly Arduino-family microcontrollers are used. On PC the client library is needed. These libraries are available for many languages like Java, Python, .NET, PHP, etc. Firmata protocol is based on MIDI messages [11].

On the Arduino side, Standard Firmata 2.3.2 version is used. The client application on PC is based on Firmata4j 2.3.3 library which is programmed in Java. The advantage of using Firmata is that another microcontroller compatible with Firmata can be used.

PN2ARDUINO extends PNEditor with many features. For Petri nets modelling, there is a capability of adding time delay to transitions and capacity for places. Also, automatic mode of firing transition was added for automatic system control purposes as only manual mode was present in PNEditor.

PN2ARDUINO brings a new communication module to PNEditor. This module communicates with the compatible microcontroller. This module consists of two parts. The first one provides the creation of connection with the microcontroller, so it sets COM port where the microcontroller is connected. The second part provides the implementation of a capability of adding Arduino components to Petri net's places and transitions. These types of Arduino components are supported: digital input and output, analog input, servo control, PWM output, message sending, custom SYSEX message [10] sending.

In Fig. 3, the use-case diagram of developed SW tool can be seen.

As it was stated, transitions and places can be associated with Arduino components. Digital and analog inputs serve as enabling conditions for transitions in Petri net. Digital and PWM outputs and messages serve as the executors of the respective actions.

TABLE I  
COMPARISON OF TWO CONCEPTS OF SYSTEM CONTROL USING PETRI NETS

Petri net logic in PC	Petri net logic in microcontroller
limited capability of real-time control	real-time control
much more computation and memory resources available	limited computation and memory resources
code in microcontroller does not need recompiling	during development repeated compiling is needed
PC must be still online	independence of control unit

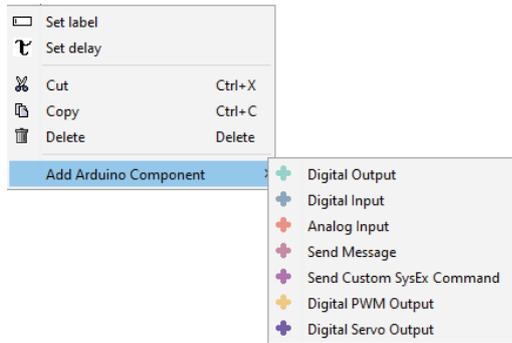


Fig. 4. PN2ARDUINO - Adding of Arduino component

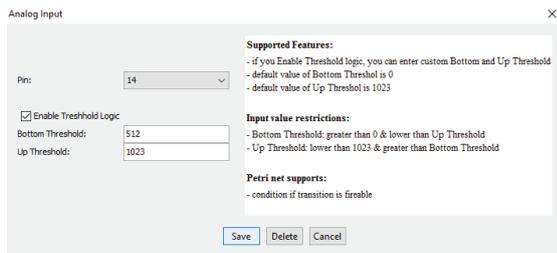


Fig. 5. PN2ARDUINO - Analog input

The interesting functionality is a capability of sending custom SYSEX messages. The user must enter SYSEX command ( $0 \times 00 - 0 \times 0F$ ) and optionally also the content of the message. The message is sent when the token comes to the place or when the transition is fired. For example, SYSEX messages are used in the proposed example of hybrid control in the last section of the paper. Here, the SYSEX message notifies the microcontroller that a different PID algorithm should be used for system control. Then PID algorithm is switched, and the controlled system remains stable.

A main window of PN2ARDUINO consists of a quick menu, main menu, canvas for Petri net modelling and log console. PN2ARDUINO supports two modes - design mode and control mode. Control mode is manual and automatic.

Firstly, it is necessary to initialize communication with Arduino (Setup board in the menu). Then it is possible to add Arduino component to the place or the transition (Fig. 4). The example of analog input can be seen in Fig. 5.

Time politics are also supported. To the transitions, it is possible to add time delay which can be deterministic or stochastic.

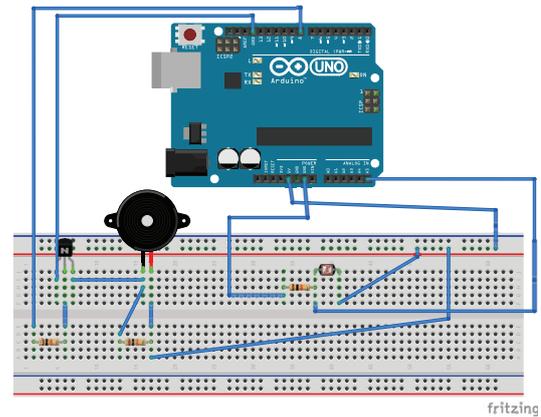


Fig. 6. The scheme of laboratory model of fire alarm

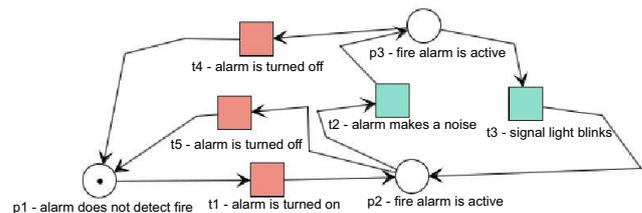


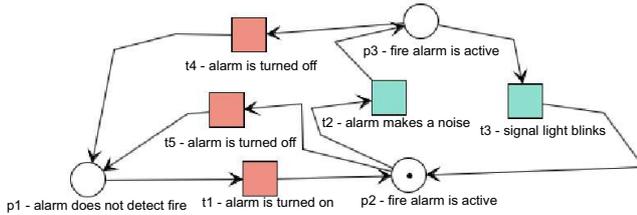
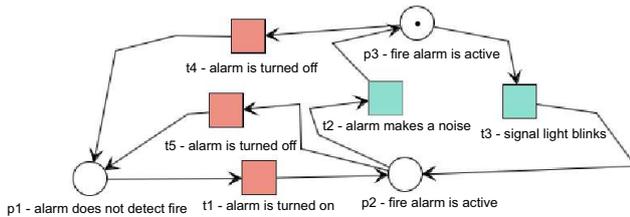
Fig. 7. PN for fire alarm (initial marking)

### III. CASE STUDY: CONTROL OF LABORATORY DISCRETE-EVENT SYSTEM

For verification of proposed software tool and method of discrete-event systems control it was necessary to design an education laboratory model of such system. A fire alarm model was built. The scheme can be seen in Fig. 6.

This model consists of an active buzzer, photo-resistor, three resistors and NPN transistor. NPN transistor is mandatory for active buzzer connection. The LED of Arduino in pin 13 is also used. Photo-resistor was used instead of the smoke sensor because of the less complicated feasibility of experiment.

Then the behaviour of the system must be defined. When the photoresistor detects an excessive lighting (it was experimentally determined as input value greater than 799 in the analog pin of Arduino Uno which resolution is from 0 to 1023) the intermittent tone of the buzzer is turned on. This tone alternates with LED lighting. When the value on the analog pin lowers below 800, these sound and light effects stop. This is repeated cyclically.

Fig. 8. PN for fire alarm ( $t_1$  is fired)Fig. 9. PN for fire alarm ( $t_2$  is fired)

Initial marking of modelled timed Petri Net interpreted for control (or sometimes called as interpreted timed Petri net) in PN2ARDUINO is shown in Fig. 7.

Places of Petri net (Fig. 7 - Fig.9) corresponds with these states:

- $p_1$  - alarm does not detect fire
- $p_2$  and  $p_3$  - alarm is active (fire was detected)

Transitions of Petri net (Fig. 7 - Fig.9) corresponds with these actions/events:

- $t_1$  - alarm is turned on
- $t_2$  - alarm makes a noise
- $t_3$  - signal light blinks
- $t_4$  and  $t_5$  - alarm is turned off

The token is in place  $p_1$  which corresponds with the state when the fire alarm is not activated because the photo-resistor does not detect light intensity threshold.

At the time when the value greater than 799 is detected on the analog pin of Arduino - the transition  $t_1$  is fired. This transition is associated with Arduino component *Analog Input* where a range of input values is set. This range determines when the transition is enabled.

Now the token is in the place  $p_2$  (Fig. 8). Transition  $t_2$  is associated with Arduino component *Digital Output* (in this case pin 8) where the buzzer is connected. This transition has also associated the function of time delay - 2 seconds. That means that transition firing (and sound effect of buzzer) lasts for 2 seconds.

Now the token is in the place  $p_3$  (Fig. 9). Transition  $t_3$  is associated with Arduino component *Digital Output* (in this case pin 13) where the build-in LED is connected. Time delay is set to 1 second. LED diode turns on for 1 second.

This process is repeated cyclically, and it is stopped when the value on the analog pin is lowered under the value 800. Then the transition  $t_4$  or  $t_5$  is fired and token moves to the place  $p_1$  when fire alarm does not detect the fire.

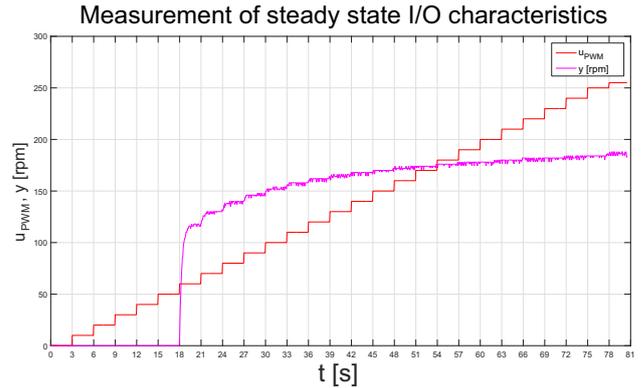


Fig. 10. Measurement of steady state I/O characteristics of DC motor

We can conclude that the ability of discrete-event control with PN2ARDUINO was successfully verified a generalized for other applications.

#### IV. CASE STUDY: CONTROL OF LABORATORY HYBRID SYSTEM

For verification of proposed software tool for hybrid systems control, it was necessary to design a laboratory model of such system. A DC motor with encoder was chosen. The encoder is used for feedback in the system because it is used for speed measurement. The actual speed of the DC is in is measured process value.

DC motor was connected to Arduino Uno using the motor shield module. Arduino motor shield is based on dual full bridge driver L298. Using the motor shield, it is possible to independently control speed and motion direction of DC motor. The encoder in this motor is of incremental type. For speed measurement, it is necessary to use hardware interruptions functionality of Arduino Uno.

The speed of the motor is set by pin described as "PWM A". When the input is set to "PWM = 255" the Arduino program shows 186 rpm which approximately corresponds with parameters stated by the manufacturer.

The next step was a measurement of steady state I/O characteristics. The input is a voltage supplied to the motor. These inputs are of size from 0V to 5V which corresponds with PWM signal from 0 to 255 (8-bit resolution). Sampling is 0.05 seconds. In Fig. 10 the process of measurement of steady state I/O characteristics is shown. The signal was filtered by 1-D median filter of 2nd order. Red line is input to the system (voltage or PWM). Output (rpm) is shown by magenta line. Steady state I/O characteristics is in Fig 11.

In the process of working points choosing it was necessary to choose points which meet the certain condition. This condition is that behaviour of the system must be close to the linear behaviour around these points. From I/O characteristics two values of input ( $u_{P_1}$  and  $u_{P_2}$ ) and output ( $y_{P_1}$  and  $y_{P_2}$ ) were chosen (these values will be our working points):

$$u_{P_1} = 80 \rightarrow y_{P_1} = 140rpm \quad (1)$$

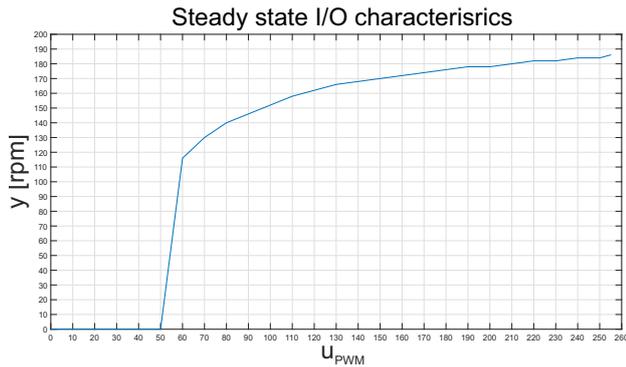


Fig. 11. Steady state I/O characteristics of DC motor

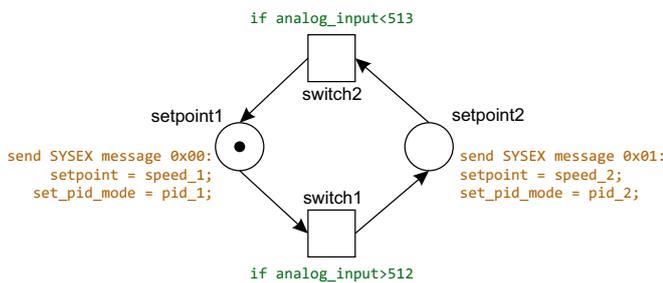


Fig. 12. Control scheme for hybrid system using PN2ARDUINO

$$u_{P_2} = 170 \rightarrow y_{P_2} = 174 \text{rpm} \quad (2)$$

From solution analysis, it is obvious that for each working point it is necessary to use a different controller. One of the solutions is an option to switch between multiple controllers according to the working point - speed (rpm) of DC motor. It is possible to use developed software module PN2ARDUINO. It is possible to switch between controllers and setpoints using SYSEX messages. Arduino and other microcontrollers that support Firmata protocol can be used. Development and verification of this software module are one of the most interesting results of presented research.

For illustration see the scheme in Fig. 12. It is an example for a demonstration of proposed control method. Assume the mentioned DC motor. We require to operate it in 2 modes (working points or rpm). For effective settlement of speed value to the setpoint, controllers with different parameters are needed (different controller for each mode). We switch between rpm using potentiometer connected to the analog input of microcontroller Arduino Uno. The switching between controllers is provided by transitions of Petri net named *switch1* and *switch2* according to the input value from potentiometer. Input from the analog pin in Arduino is represented by value between 0 and 1023. As the threshold, a half value was used (512). In the moment when the token in Petri net is moved to the place named *setpoint1* or *setpoint2*, a SYSEX message is sent. This message ensures the execution of user defined program code on the Arduino side. In this case, the control algorithm is executed. An algorithm (PID controller)

for continuous control is independent of Firmata messaging, so it provides real-time control.

The case study of hybrid systems control proposed a basic example. Researchers in the field of hybrid control design can use it for different and more complicated scenarios.

## V. CONCLUSION

The paper presents the new software tool named PN2ARDUINO which extends PNEditor with the capability of communication with microcontrollers that supports Firmata protocol. Then it is possible to control discrete-event and hybrid systems using timed interpreted Petri nets with developed software tool. This tool uses the control paradigm when the microcontroller has implemented only the communication protocol. Petri net's control logic is stored in the computer which communicates with the microcontroller and sends control orders. The next research will focus on the concept of control with Petri nets where control logic will be directly implemented on the microcontroller.

## ACKNOWLEDGMENT

This work has been supported by the Cultural and Educational Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic, KEGA 030STU-4/2017 and KEGA 038STU-4/2018, by the Scientific Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic under the grant VEGA 1/0733/16, and by the Young researchers support program, project No. 1328 – KVPRI (Quality Control of Production Processes with Augmented Reality in Industry 4.0) and project No. 1327 - VTOVI (Virtual Training of Production Operators for Industry 4.0).

## REFERENCES

- [1] H. Matsuno, A. Doi, R. Drath, and S. Miyano, "Genomic object net: object oriented representation of biological systems," *Genome Informatics Series*, pp. 229–230, 2000.
- [2] M. A. Drighiciu and G. Manolea, "Application des reseaux de petri hybrides a l'etude des systemes de production a haute cadence," 2010.
- [3] P. J. Mostermans, M. Ottery, and H. Elmqvist, "Modeling petri nets as local constraint equations for hybrid systems using modelica," *retrieved online at http://citeseer.ist.psu.edu/359408.html*, 1998.
- [4] S. Fabricius and E. Badreddin, "Modelica library for hybrid simulation of mass flow in process plants," in *Proceedings of the 2nd International Modelica Conference*, Oberpfaffenhofen, Germany. Citeseer, 2002, pp. 225–234.
- [5] S. Proß, B. Bachmann, and A. Stadtholz, "A petri net library for modeling hybrid systems in openmodelica," in *submitted (Modelica Conference 2009)*, 2009.
- [6] S. Proß and B. Bachmann, "Pnlib-an advanced petri net library for hybrid process modeling," in *Modelica Conference*, 2012.
- [7] M. Riesz, M. Seckár, and G. Juhás, "Petriflow: A petri net based framework for modelling and control of workflow processes," in *ACSD/Petri Nets Workshops*. Citeseer, 2010, pp. 191–205.
- [8] A. Ceseckova, "Control of laboratory discrete event systems (in slovak)," Master's thesis, Slovak University of Technology in Bratislava, 2016.
- [9] E. Kucera, "Modelling and control of hybrid systems using high-level petri nets (in slovak)," Ph.D. dissertation, Slovak University of Technology in Bratislava, 2016.
- [10] H.-C. Steiner, "Firmata: Towards making microcontrollers act like extensions of the computer," in *NIME*, 2009, pp. 125–130.
- [11] M. Association. (2016) Summary of midi messages. [Online]. Available: <https://www.midi.org/specifications/item/table-1-summary-of-midi-message>



# Use of Holographic Technology in Online Experimentation

Jakub Matisák

Slovak University of  
Technology, Faculty of Electrical  
Engineering and Information  
Technology, Ilkovičova 3, 812 19  
Bratislava, Slovakia  
Email: jakub.matisak@stuba.sk

Matej Rábek

Slovak University of  
Technology, Faculty of Electrical  
Engineering and Information  
Technology, Ilkovičova 3, 812 19  
Bratislava, Slovakia  
Email: matej.rabek@stuba.sk

Katarína Žáková

Slovak University of  
Technology, Faculty of Electrical  
Engineering and Information  
Technology, Ilkovičova 3, 812 19  
Bratislava, Slovakia  
Email: katarina.zakova@stuba.sk

**Abstract** - The paper deals with a web application that allows simulating 3D model of mechatronic system in holographic devices. Purpose of this application is to bring new perspective to interactive education and simplify the process of studying. Background of an application is driven by Scilab Xcos simulation software communicating via web service, which provides data. The resulting animation is displayed on a holographic device, which allows visualization. The displayed system is a 3D model of the mechatronic experiment, which represents digital model of real device. Accurate movement of experiment is obtained by linking data from Scilab Xcos with 3D model. The 3D model visualization should help with easier understanding of the subject matter.

**Keywords** - mechatronic system, simulation, animation, holographic technology

## I. INTRODUCTION

THE goal of each educational institution should be to improve and innovate the educational process. New methods of educating students should be formed to make the process easier, such as in [1]. Also, it is necessary to increase its efficiency. The ideal outcome is to bring education and research activities together, creating innovations that support the industry [2]. The latest trend in education is to bring an application that gives a better understanding of the issue. Simplifying device designs, understanding technical specifications, facilitating device prototyping, or even making manufacturing process cheaper are just a few of the many different uses of 3D hardware digitalization [3]. There are many three-dimensional environments around the world that try to incorporate, work, and simulate knowledge from different areas [4]. Study says that at laboratory sessions 58% students agreed that methodologies like simulations, demonstrations and virtual labs make them more comfortable in lab sessions [5]. Nowadays, we can observe the trend of digitizing [6] and simulating equipment in almost every working segment. It allows us to face real situations before they happen, to learn from them, see issues from another perspective, respond to them much faster and, finally, to save costs. Lately many institutions have specialized in virtual and augmented reality, like in [7] and brought attention to it. However, our project wants to focus on an area that is not so widespread. The aim of our work is to show students another angle of learning. To do so, we used holographic technology.

Optical holography for recording three-dimensional scenes can be traced back to the early sixties. Since then, the art of holography has been applied in many areas, primarily as a tool for 3D imaging, processing, and display [8]. Study in [9] says that 45.5% of teachers believe that hologram technology would have affect in the field of teaching. The use of holographic technology could be used in various areas of life. The first example is using it in car, which is published in [10]. The authors attempt to present a holographic display, that would help reduce the time when drivers were guided to the dashboard. Hologram would be projected onto the front glass, so time of inattention would be reduced. Another example is in medicine. There is a possibility of displaying real heart beating on a model of heart in four-sided hologram pyramid, which authors published in [11].

The aim of this paper is to help students with understanding of the subject matter dealing, for instance with the basics of automatic control. Our system can simply help to visualize the behavior of mechatronic experiment parts as 3D digital model in holographic device.

## II. STATE OF TECHNOLOGY

The “Hologram” word refers to a three-dimensional picture made by laser light reflected onto a photographic substance without the use of a camera [12]. Hologram device could be used to play video, represent some system behavior, show object models, etc. We know many varieties of holograms, and there are variable ways of classifying them. For our purpose, we can divide them into three main types: reflection, transmission and hybrid holograms.

### A. Reflection holograms

The reflection hologram (Fig. 1) is the most common type of the hologram. They can be seen in galleries and in presentation places. This hologram is formed when the reference beam and the object beam are incident on opposite sides of the holographic surface. They interfere and record an image. To reconstruct the image, a point source of white light illuminates the hologram from the proper angle, and the viewer looks at it from the same side as the light source. Reflection holograms require the simplest setup and are visible without laser light [13].

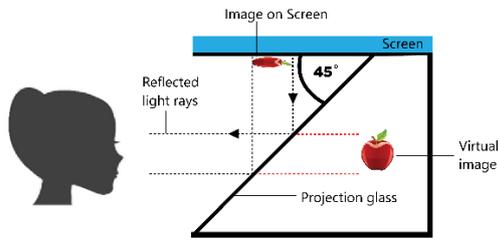


Fig. 1 How reflection hologram works

### B. Transmission holograms

Transmission holograms (Fig. 2) are also known as Laser-Transmission Holograms. This type is created when the reference beam and the object beam are incident on the same side of the holographic surface. They are viewed by shining a spread-out laser light through the emulsion side of the hologram at the same angle the hologram was recorded at, with the viewer looking on from the opposite side. The light is transmitted from behind the hologram device to the side of the observer [13]. Image which is displayed can be very precise.

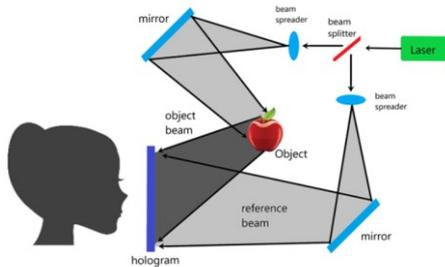


Fig. 2 How transmission hologram works

Materials, methods, and processes used to make transmission hologram are the same as reflection holograms.

### C. Hybrid holograms

As hybrid hologram could be considered a combination of transmission and reflection hologram. Hybrid hologram can be specified as multichannel holograms, holographic interferometry, integral holograms, embossed holograms, and computer-generated holograms. For example, embossed holograms are used for authenticity applications such as security hologram stickers, passports or credit cards. Computer-generated holograms are used to make optical elements, for scanning, splitting, in general for controlling laser light, example can be CD player [14]. These types of hologram are not relevant for our work, because of their technology, so we will not pay attention to them.

## III. PRODUCT RESEARCH

Looking on the market the most advanced holographic device suitable for our purpose can be considered Microsoft HoloLens. HoloLens is a pair of mixed reality smart glasses, which is a holographic computer built into a headset that lets you see, hear, and interact with holograms within an environment such as a living room or an office space [15]. It is a wearable device that permits to look at holograms that are

connected to the world and interact with them using gestures, voice commands and gaze. For instance, one of its advantages is plugin for Unity engine. However, due to its higher price (3500\$), it is unlikely that will be massively adapted soon.

Next, the Realfiction company offers a range of holographic devices. These devices have different sizes and differences in the number of display areas. They also have higher prices (2000\$ - 10 000\$). In our project we approached to use the Realfiction Dreamoc HD3.2. Device has three-sided view, HDMI port, RJ45 port 23" screen and built-in loudspeakers. More details can be found in [20]. We chose it because this device has connectivity advantage - HDMI.

For everyday use it is possible to buy quite cheap holographic devices and use it with smartphones or tablets for example from company Holho. Price can be from 40\$ to 160\$. Disadvantage of these devices is that only video can be played, so when we want to control behavior of real-time experiment it is not possible to change anything in process without connection to another device via cable.

## IV. APPLICATION

We had following requirements for the application:

- to make realistic view of 3D mechatronic experiment,
- to control the behavior of the 3D model using parameters entered by the user.

### A. Hardware

As it was said before, it's necessary to have hardware to generate holographic image. This hardware consists of two main parts. The device from the front side is shown in Fig.3.



Fig. 3 Hologram device from the front side

The first part is the image-emitting screen. A conventional computer screen could be used, but the resolution of screen increases the quality of the image. Our device has a screen up and emits the image from top to bottom. The second part is a projection glass with a semi-permeable layer, which is placed at a 45-degree angle below the screen. Therefore, the screen provides an image that is reflected on the glass. By placing the glass at this specific angle, the image is presented as if it is behind the glass, which creates virtual image of the represented object.

### B. Software and system architecture

Application with minimal requirements and possibility of massive adoption in future led us to use standard web technologies (HTML, CSS, JavaScript and Three.js). It is necessary to realize that the display area is in the hologram, so it is not possible to change the control parameters from there. To do so, we needed to use two browser windows. The

first one is opened in hologram as a view and the second one in computer screen as a control window. It is allowed via multi-screen mode. Of course, whole system is more complicated, and its architecture is shown in Fig.4.

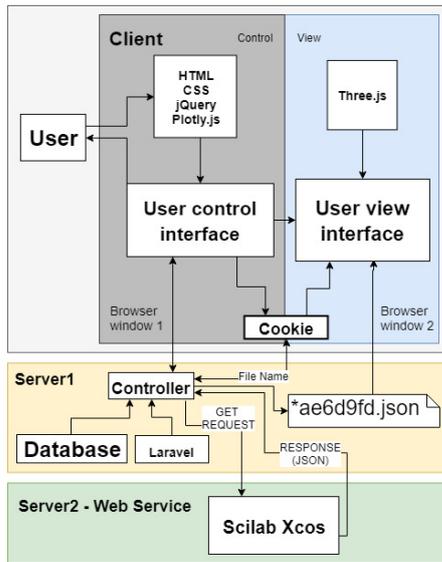


Fig. 4 System architecture

Back-end side of the application is managed from the server using a Laravel framework. When control interface is opened, data for specific experiment are automatically downloaded from MySQL database and a HTML form is generated. This form is generic, to ensure the possibility of adding new experiments in the future. The realistic movement visualization can be achieved thanks to numerical data received from Scilab Xcos simulation environment where the experiment is represented as a block diagram. Scilab allows to simulate a block scheme by accessing it through a terminal. Module with Scilab has its own interface, works as standalone application and it is used as a web service. It is necessary upload block scheme to the server before the first simulation. Then, it is used by an authorized program, accessed via URL. The main problem of showing data in 3D model is to send them from control to view interface or to change the simulation parameters during the visualization. Since each client window works separately, it is necessary to inform view interface about parameter changes in control interface. Solution we used is simple, when control view is opened, random SHA code is generated and saved as cookie value. This cookie is used then as a file name for data from simulation. Data are sent as a HTTP response from service and always saved to this specific file during entire session. User view interface gets these data and starts to render movement of experiment.

C. Web Service

There are several simulation programs on the market, like Matlab, Scilab Xcos or Octave. In our case, we have chosen Scilab Xcos, which is an open source distribution modeling and simulation software for numerical computation. The choice was done from several reasons. Matlab as the most

used software requires license and not everyone have access to it. Therefore, it was not suitable for this implementation. Octave from the open source category does not provide a block diagram option, which is suitable for easier creation of controllers. There exist also other programs that have a graphical editor for building block diagrams and have also appropriate numerical methods to solve differential equations. However, Scilab Xcos is the closest open source option to Matlab.

Secondly, since we want the system to be modular, it is very useful to have an Application Programming Interface (API) that allows us to use this simulation software as a web service. The advantage of the simulation environment that is located on the server and works as a web service is that everyone who requires to retrieve the data needs to know only the URL and how to access the software through the appropriate API. We have already had API for Scilab Xcos at the time of implementation, so we decided to use it. Since the entire data processing module is running as a web service, it can be easily replaced by other software. In this way the application can be expanded by a different simulation environment in the future.

V. ONLINE EXPERIMENT

In Fig.7 a model of the Furuta pendulum for our holographic device is depicted. The holographic device we used renders three-side projection, so it is necessary to have the model displayed three times and rotated by ninety degrees. The one on the bottom is then shown on the front side of the device and the other two on each side.

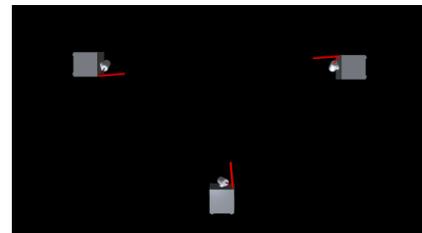


Fig. 5 Furuta pendulum view in system

Firstly, it is necessary to open view interface (Fig.5) in holographic device.

The control interface is a web form titled 'Controllers'. It contains several input fields and a 'Start simulation' button. The parameters are as follows:

Program	K1	Required angle
Scilab	-3.5443	60
Controller	K2	Sampling period
State	-2.1498	0.01
	K3	Animation time
	-20.0816	5
	K4	
	-4.6941	

Start simulation

Fig. 6 Control interface for input data

Secondly, data are required. They are coming from control interface where the block scheme is simulated. In our case the control structure includes model of Furuta pendulum and the state space controller. Its parameters (i.e. gains  $K1$ ,  $K2$ ,  $K3$ ,  $K4$ ) can be defined via user interface shown in Fig.6. User can specify also *required angle*, *sampling period* and *animation*

time. After sending the request to the server (by clicking on the button *Start simulation*), the data will arrive within seconds and the model simulation will automatically start. The input parameters can be altered by sending a new request. The process is designed to change movement automatically. In Fig. 7 is shown holographic device Dreamoc HD3.2 with connected computer via HDMI.



Fig. 6 Control interface in computer screen and Furuta pendulum model in holographic device

## VI. CONCLUSION

The paper describes web application for holographic devices, developed mostly for educational purposes. It can be used as a tool for interactive teaching in subjects dealing with automatic control theory. With this technology, for students it is simpler to imagine mechatronic systems behavior and that should be goal of every educational institution. The application acts as simulation software for 3D objects that helps students to understand how to operate with specific devices and systems. The application uses Scilab Xcos as simulation environment for computing block schemas. Achieved simulation data allow to see realistic movement of experiment 3D model. The 3D model, which is shown as holographic mechatronic experiment projection of the specific device. Holographic device that was used to show model and the behavior of the experiment, cannot be considered as a solution that could be used massively, due to its higher price. The tool can be used as a display device for teachers during lessons. However, the application can be simply modified to a device that is not subject of a high price and this version could be deployed for every student.

As a future work, authors would like to extend the application by directly embedding it into the online laboratory portal, which deals with real-time control of experiments. This portal is currently in the implementation phase. Also, they would like to extend the application to use it through augmented reality, using Google ARCore platform.

## ACKNOWLEDGMENT

The work presented in this paper has been supported by the Cultural and Educational Grant Agency of the Ministry of

Education, Science, Research and Sport of the Slovak Republic, KEGA 030STU-4/2017 and by the Tatra Banka Foundation within the grant program E-talent, project No. 2018et016 (Holographic technology and augmented reality in online experimentation). Authors would like to thank to all contributors for help with implementation.

## REFERENCES

- [1] B. Kraut a J. Jeknić, „Improving education experience with augmented reality (AR),“ rev. 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 2015. doi:10.1109/MIPRO.2015.7160372
- [2] J. Autiosalo, „Platform for industrial internet and digital twin focused education, research, and innovation: Ilmatar the overhead crane,“ rev. 2018 IEEE 4th World Forum on Internet of Things (WF-IoT), Singapore, Singapore, 2018. doi:10.1109/WF-IoT.2018.8355217
- [3] K. Žáková a M. Hók, „Interactive three dimensional presentation of Segway laboratory model,“ rev. 2016 International Conference on Emerging eLearning Technologies and Applications (ICETA), Vysoké Tatry, Slovakia, 2016. doi:10.1109/ICETA.2016.7802064
- [4] J. Lebieź a M. Szwoch, „Virtual sightseeing in Immersive 3D Visualization Lab,“ rev. 2016 Federated Conference on Computer Science and Information Systems (FedCSIS), Gdansk, Poland, 2016.
- [5] M. Deepak, T. Sandeep, B. Shruti, B. Manish a S. Sneha, „An analysis to find effective teaching methodology in engineering education,“ Jaipur, India, 2013. doi:10.1109/MITE.2013.6756331
- [6] B.-M. Block, „Digitalization in engineering education research and practice,“ rev. 2018 IEEE Global Engineering Education Conference (EDUCON), Tenerife, Spain, 2018. doi:10.1109/EDUCON.2018.8363342
- [7] E. Kucera, O. Haffner a R. Leskovský, „Interactive and virtual/mixed reality applications for mechatronics education developed in unity engine,“ rev. 2018 Cybernetics & Informatics (K&I), Lazy pod Makytou, Slovakia, 2018. doi: 10.1109/CYBERI.2018.8337533
- [8] Peter Wai Ming Tsang; Ting-Chung Poon, „Review on the State-of-the-Art Technologies for Acquisition and Display of Digital Holograms,“ IEEE Transactions on Industrial Informatics, zv. Volume: 12, %1. vyd.Issue: 3, pp. 886 - 901, June 2016. doi:10.1109/TII.2016.2550535
- [9] H. Ghuloum, „3D Hologram Technology in Learning Environment,“ rev. Proceedings of Informing Science & IT Education Conference (InSITE) 2010, Cassino, Italy, 2010.
- [10] W. Wang, X. Zhu, K. Chan a P. Tsang, „Digital Holographic System for Automotive Augmented Reality Head-Up-Display,“ rev. 2018 IEEE 27th International Symposium on Industrial Electronics (ISIE), Cairns, QLD, Australia, 2018. doi:10.1109/ISIE.2018.8433601
- [11] T. Thap, Y. Nam, H. Chung, J. Lee, „Simplified 3D Hologram Heart Activity Monitoring Using a Smartphone,“ rev. 2015 9th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, Blumenau, Brazil, 2015. doi:10.1109/IMIS.2015.87
- [12] A. Abdullah H a K. Faten F., „The first design of a smart hologram for teaching,“ rev. 2018 Advances in Science and Engineering Technology International Conferences (ASET), Abu Dhabi, United Arab Emirates, 2018. doi:10.1109/ICASET.2018.8376931
- [13] K. Fey, „Creation of Simple Holograms with Single Beam Setup,“ August 2000. [Online]. Available: [https://laser.physics.sunysb.edu/\\_karl/webreport/index.html](https://laser.physics.sunysb.edu/_karl/webreport/index.html). [Cit. 6 May 2019].
- [14] A. Jeong; T. H. Jeong, „What Are the Main Types of Holograms?,“ Integraf LLC, 2019. [Online]. Available: <https://www.integraf.com/resources/articles/a-main-types-of-hologram-s>. [Cit. 6 May 2019].
- [15] J. Roberts, „What is HoloLens? Microsoft’s holographic headset explained,“ Trusted Reviews, 30 March 2016. [Online]. Available: <https://www.trustedreviews.com/opinion/hololens-release-date-news-and-price-2922378>. [Cit. 6 May 2019].
- [16] Realfiction, „Dreamoc Fact-sheet,“ [Online]. Available: [https://www.dropbox.com/s/x7n1gvkh2yfmyf3/dreamochd\\_factsheet\\_version-1.pdf?dl=0](https://www.dropbox.com/s/x7n1gvkh2yfmyf3/dreamochd_factsheet_version-1.pdf?dl=0). [Cit. 1 May 2019].

# Proposal of Mechatronic Devices Control using Mixed Reality

Erich Stark, Erik Kučera, Peter Drahoš, Oto Haffner  
Slovak University of Technology in Bratislava  
Bratislava, Slovakia  
Email: erich.stark@stuba.sk

**Abstract**—The Internet of Things and mixed reality are now among the most important areas in research or in practice. The aim of this paper is to propose an appropriate way of connection of these two areas, where is possible to control and monitor mechatronic devices using a mobile device with augmented/mixed reality support. The main task will be to explore these options in the area and implement this solution as prototype. The proposed methodology for control and diagnostics of mechatronic devices is modern as it combines hardware management, a Unity3D engine for mixed reality development, and communication within the Internet of Things network.

## I. INTRODUCTION

CURRENTLY, computer networks are no longer just for connecting conventional computers like they once were. Their purpose gained a new dimension when mobile devices and embedded systems began to connect to these networks. At present, these boundaries are shifted to the level of connection of individual sensors, various household appliances, and even autonomous cars to the network [1]. This expansion of connected devices has also happened because of the rise of microcomputers like Raspberry Pi, DragonBoard, and similar prototyping solutions. At the same time, people begin to realize the value of data that these sensors generate. They can help us streamline processes in industry and services or make life easier with smart home solutions. As a result, the emergence of new types of networks such as Internet of Things (IoT) are needed. The concept of IoT can be found at almost every conference in the field of information and communication technologies or in scientific articles [10], [11], [12]. The Gartner company makes regular analyzes and research into the use of various technologies. Earlier in 2017, an analysis was made that states that IoT will have up to 20.5 billion connected devices in 2020 [2]. These paradigms would not take place without the development of new networks, data transmission protocols and the necessary software tools. At present, IoT devices are controlled by console, web, or mobile applications. Using these conventional methods of controlling IoT devices in a small room can be quite simple. Because the list of devices is on one screen, we can see and set properties almost instantly. But if there are multiple rooms or buildings, the segmentation of these devices may be totally unclear and cumbersome. Here is the opportunity to use current trends and modern technologies in the field of virtual, augmented and mixed reality. These technologies are able to put digital objects into the real world. Their convenience lies

in the fact that objects from the real world are enriched with information relevant to the given object that one is looking at. This camera stream processing is real-time. Mixed reality can now be developed and tracked with compatible headset - such as Microsoft HoloLens or compatible mobile devices (smartphones and tablets) from both leaders in the segment - Google Android and Apple iOS. The implementation of the proposed project involves the use of mobile devices for their wide availability - whether for household or industry. Compatible headsets are currently not suitable for this purpose, as businesses (especially small and medium-sized ones) are often unwilling to invest in these headsets. The proposed methodology for controlling and diagnosing IoT devices is modern as it combines hardware management, a 3D engine for mixed reality development, and communication within the Internet of Things network - all areas of mechatronics. The proposed solution is unique and will contribute to the scientific field of mechatronics.

## II. COMPUTER GENERATED REALITY

### A. Virtual Reality

Virtual reality (VR) is a term that is mentioned in various areas, not only in information and communication technologies. Films like Matrix have brought virtual reality from the sci-fi world to the human mind. Examples of virtual and extended reality are becoming more and more real-life, from military air simulators to simple smartphone applications. Everyone can have their own idea of virtual reality, so it is necessary to introduce a suitable definition.

Virtual reality consists of an interactive computer simulation that senses the state of the user and replaces or extends sensory feedback information to one or more senses in such a way that the user gets a feeling of being immersed in the simulation (virtual environment).

### B. Augmented Reality

Augmented reality (AR) is an overlapping of content in the real world, but this content is not embedded or part of it. The content of the real world is not capable of responding to computer-generated content [3]. Augmented reality is therefore a live, direct or indirect view of a real world that is complemented by computer generated (CG) elements such as audio, video, graphics, or GPS data. Augmented reality is a layer of content above the real world, and this content is not

anchored to this world or its part. As has been said, elements of the real world and CG content can not react with each other.

The purpose of Augmented reality is to improve user perception and improve its effectiveness through additional information. The user retains awareness of the real world, but in an ideal extended reality it would not be able to recognize the difference between information from the real world and the virtual world.



Fig. 1. Augmented reality example

### C. Mixed reality

Mixed reality (MR) is an overlap of the real world with synthetic content that is embedded in it and interacts with the real world. The key feature of MR is that real-world synthetic content and content can respond in real time to one another. Mixed reality is thus a combination of the real world and the virtual world, creating a new environment and visualization where physical and digital objects coexist and interact with each other in real time. Mixed reality is the layer of artificial (digital) content in a real world that is anchored and interacts with the real world. An important fact is that, in the case of mixed reality, advanced mapping of the environment is required for the placement of additional CG elements.

If information is to be successfully combined, virtual objects must act physically in a suitable way. If a real and virtual object collision occurs, both must respond appropriately. In addition, virtual objects must overlay the view of real objects and also shadow on them. All this can only be achieved by a precise model of real and virtual environments.

The first hardware for mixed reality, currently the most advanced device of the segment, is Microsoft HoloLens. The problem is still a relatively high price, but there is also an emulator for development.

Based on the information we have mentioned, mixed reality seems to be the most exciting. However, it is possible to imagine the future in which synthetic content will be able to react in some way and even communicate with the real world [3].

### III. INTERNET OF THINGS

Internet of Things is currently a very widespread term in the field of modern information and communication technologies.



Fig. 2. Mixed reality example

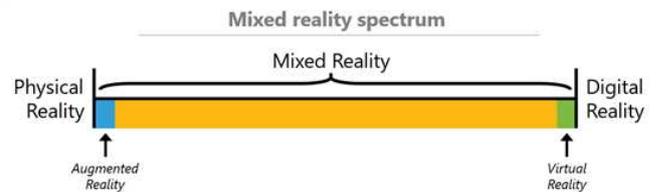


Fig. 3. Mixed reality spectrum

This issue is the subject of various debates as its deployment in industry and services brings about more effective action, but it also raises various issues, such as safety. Thus, IoT concerns almost all fields of human activity [4].

In general, IoT can be defined as a set of physical objects (or things) embedded in electronics, software, sensors and connected devices that are connected together in the network to allow data exchange with other interconnected devices to achieve higher value and more services for users. These IoT devices create a linked network in which each is uniquely identifiable with a unique IP address and capable of communicating with existing network infrastructures.

### IV. RELATED WORK

In the paper [5], the authors presented the Augmented Things concept, where computer objects contain all the information needed to track and expand the information required by AR applications. This allows the user to connect to them, retrieve information using their mobile device, and get expanding information like, for example, maintenance, device, or usage information. The authors have created also a simple 3D framework that allows you to track objects using high quality 3D high resolution scans.

Phillipe Lewicki has attempted to create a demonstration program to help control the Philips Hue light bulb using the Microsoft HoloLens device as seen in Fig. 5. He realized that today's solutions allow you to control the bulbs using the mobile application they need to open, to find a particular room, and then a particular bulb. Often, such applications are limited because smart bulbs contain more features than just turn on / off.

Thanks to the HoloLens on the head, it was only possible to look at the light and turn it into a simple gesture or change

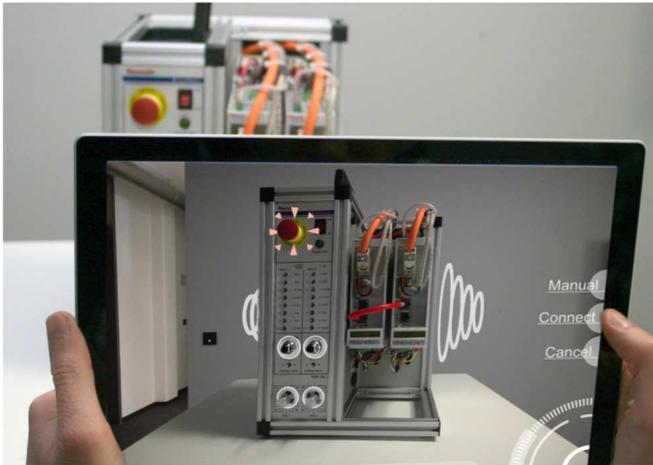


Fig. 4. Augmented Things objects contain and share their AR information



Fig. 5. Application for HoloLens, enabling color adjustment of light

the color of the light. It was faster than a wall switch [6]. Fig. 6 shows Proof of Concept (PoC) by designer Ian Sterling and software engineer Swaroop Pala. Their concept shows how smart devices could be controlled by gestures. The task of this project was to provide a 3D user interface with Android Music Player and Arduino light fan [7].

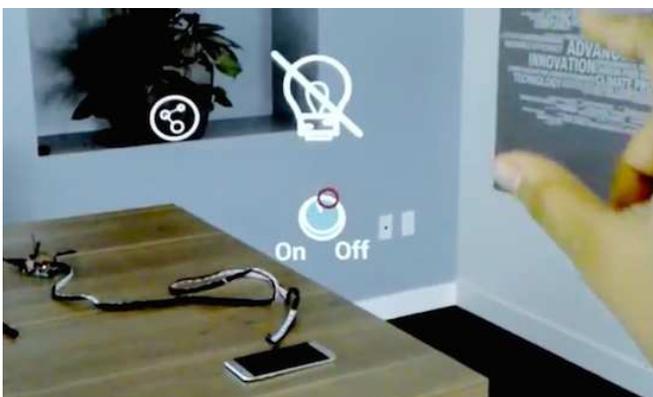


Fig. 6. Light control using HoloLens and Arduino devices

### V. SYSTEM PROPOSAL

The diagram of the proposed system can be seen in the Fig. 7.

System description:

- 1) At the beginning of the system is mixed reality device which is able to analyze data stream from camera and detects QR code.
- 2) Application can connect to identical object in the cloud.
- 3) Data from the device sensors are sent to the cloud.
- 4) Mixed reality application gets information about device and shows tailored user interface.
- 5) The user can interact with that device using mixed reality experience.
- 6) It is possible to send some control commands to the cloud.

This system can be decoupled into several components described in subsections below.

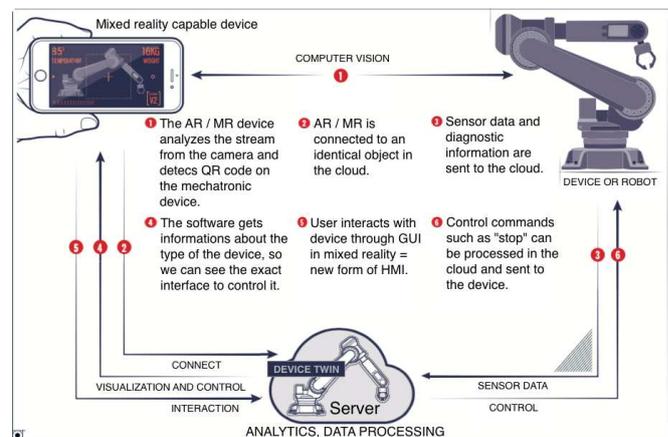


Fig. 7. Diagram of system proposal

#### A. Camera Device with Mixed Reality Support and Computer Vision Tools

The best option for mixed reality experiences nowadays is smartphone when comes to price or availability for masses. Basically, there are two options: Android system with ARCore SDK or iOS with ARKit SDK support. ARKit SDK was chosen because of greater support of functions needed for this project. For example, like persistent content or 3D object recognition in recently announced ARKit 2.

At first it has to be created software which can analyze video stream from smartphone camera and detects physical objects. ARKit can be used with support of framework Vuforia inside Unity 3D editor, which helps a lot with software development. The main features of Vuforia SDK is *Multi Target detection*, *User Defined Targets* or *Cloud Recognition*.

#### B. Software Platform for IoT

OPC is currently the most advanced standardized data exchange process for automation technology. It allows the collection and transmission of data in a unified form from

various devices, control systems and applications throughout the organization. The design of this standard allows mapping almost all industrial data into the OPC data structure. OPC UA is an enhanced version of the OPC standard that has a unified architecture that makes it a platform-independent protocol. In addition, it has built-in security mechanisms and applications are fully scalable from microcontrollers to corporate servers.

### C. IoT Prototyping Hardware Kit

There were many options for prototyping devices from single hardware to complex IoT kits. It was not easy to find IoT kit which meet system requirements the most.

BigClown is a modular hardware and software system that allows to prototype and build real-world telemetry, automation and other applications including IoT. BigClown can be imagined as a set of components with a single interface that can be connected together depends on application needs.

The core of each device is the so-called Core Module. It is powered by a single core CPU with the Cortex-M0+, specifically STM32L083CZ. This chip was chosen for a number of reasons: it is proven and used ARM CPU from STM32 series, has a very low consumption (which is important for powering the nodes from battery), has integrated USB with ROM bootloader, enough number of interfaces, (Flash, RAM and EEPROM), and above all, it has two cryptographic components: TRNG (True Random Number Generator) and AES-128 computing accelerator [8].



Fig. 8. Whole BigClown ecosystem [8]

## VI. CONCLUSION

The upcoming trend Internet of Things has an impact not only on applications for various services, households and intelligent buildings, but also significant impact on industry and industrial production. The application of IoT principles in industry is called Industrial Internet of Things (IIoT), where in this case instead of interconnected devices is used individual machine parts or their sensors and actuators, as well as sensors and actuators for HVAC (Heating, ventilation, and air conditioning) security. Device interconnection should be wireless in particular and should bring new interaction capabilities not only between systems, but also bring new capabilities to control, track and secure advanced services.

This proposal of interconnection IoT and mixed reality can bring new form of Human Machine Interface which can save time for users or companies.

The future work will focus on developing such a solution with technologies mentioned before.

### ACKNOWLEDGMENT

This work has been supported by the Cultural and Educational Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic, KEGA 030STU-4/2017 and KEGA 038STU-4/2018, and by the Young researchers support program, project No. 1324 – VZRI4 (Virtual and Mixed Reality for Industry 4.0) and project No. 1325 - ODIZPZR (Monitoring and Diagnostics of IoT Devices using Mixed Reality).

### REFERENCES

- [1] Hammar, Sven. *Connected cars: Driving the Internet of Things revolution: How to run an IoT enabled business* [online]. IoT Now, 2017 [cit. 2018-03-23]. Available at: <https://www.iot-now.com/2017/04/03/60270-connected-cars-driving-internet-things-revolution/>
- [2] Meulen, Rob van der. *Gartner Says 8.4 Billion Connected "Things" Will Be in Use in 2017, Up 31 Percent From 2016: IoT Units Installed Base by Category* [online]. Gartner, 2017 [cit. 2018-03-23]. Available at: <https://www.gartner.com/newsroom/id/3598917>
- [3] *Virtual Reality: VR, AR, MR* [online]. The Foundry [cit. 2018-04-22]. Available at: <https://www.foundry.com/industries/virtual-reality/vr-mr-ar-confused>
- [4] Pohanka, Pavel. *Internet veci* [online]. i2ot, 2017 [cit. 2018-03-25]. Available at: <http://i2ot.eu/internet-of-things/>.
- [5] Rambach, Jason et al. *Augmented Things: Enhancing AR Applications leveraging the Internet of Things and Universal 3D Object Tracking*. In: IEEE International Conference on Industrial Technology (ICIT). 2017, zv. 22, s. 25.
- [6] Lewicki, Philippe. *Controlling lights with the Hololens and Internet of Things* [online]. htmfufusion, 2016 [cit. 2018-04-03]. Available at: <http://blog.htmfufusion.com/controlling-lights-with-the-hololens-and-internet-of-thingsatch-one-of-philippes-appearances-in-june/>
- [7] Sterling, Ian a PAL, Swaroop. *Control with your smart devices by staring and gesturing* [online]. Arduino, 2016 [cit. 2018-04-03]. Available at: <https://blog.arduino.cc/2016/07/26/control-with-your-smart-devices-by-staring-and-gesturing/>
- [8] Maly, Martin. *BigClown: IoT jako modulární stavebnice* [online]. Root.cz [cit. 2018-04-22]. Available at: <https://www.root.cz/clanky/bigclown-iot-jako-modularni-stavebnice/>
- [9] T. Patys, K. Murawski, A. Arciuch and A. Walczak, "Optical driving for a computer system with augmented reality features," 2017 Federated Conference on Computer Science and Information Systems (FedCSIS), Prague, 2017, pp. 657-662.
- [10] M. P. Loria, M. Toja, V. Carchiolo and M. Malgeri, "An efficient real-time architecture for collecting IoT data," 2017 Federated Conference on Computer Science and Information Systems (FedCSIS), Prague, 2017, pp. 1157-1166.
- [11] R. Falkenberg et al., "PhyNetLab: An IoT-based warehouse testbed," 2017 Federated Conference on Computer Science and Information Systems (FedCSIS), Prague, 2017, pp. 1051-1055.
- [12] J. Mocnej, T. Lojka and I. Zolotová, "Using information entropy in smart sensors for decentralized data acquisition architecture," 2016 IEEE 14th International Symposium on Applied Machine Intelligence and Informatics (SAMI), Herlany, 2016, pp. 47-50.

# The Effects of Augmented Training Dataset on Performance of Convolutional Neural Networks in Face Recognition System

Mehmet Ali Kutlugün  
Computer Science and Engineering  
Istanbul Sabahattin Zaim University  
Istanbul, Turkey  
mehmet.kutlugun@std.izu.edu.tr

Asst.Prof. Yahya Şirin  
Computer Science and Engineering  
Istanbul Sabahattin Zaim University  
Istanbul, Turkey  
yahya.sirin@izu.edu.tr

MehmetAli Karakaya  
Management Information Systems  
Anadolu University  
Eskişehir, Turkey  
mehmetalikarakaya@aof.anadolu.edu.tr

**Abstract**— Nowadays, deep learning methods have been used in many areas such as big data analysis, speech and image processing with the increasing processing power and the development of graphics processors. In particular, face recognition systems have become one of the most important research topics in biometry. Light direction, reflection, emotional and physical changes in facial expression are the main factors in face recognition systems that make recognition difficult. Training of the system with the available data in small data sets is an important factor that negatively affects the performance. The Convolutional Neural Network (CNN) model is a deep learning architecture used for large amounts of training data. In this study, a small number of employee images set of a small-scale company has been increased by applying different filters. In addition, it has been tried to determine which data augmentation options have more effect on face recognition. Thus, non-real-time face recognition has been performed by training with new augmented dataset of each picture with many features.

**Keywords**— deep learning; convolutional neural networks; image processing; face recognition; data augmentation

## I. INTRODUCTION

### A. Deep Learning

Deep learning is a series of machine learning algorithms that try to model high-level abstractions of data with model architectures from multiple nonlinear transformations. Deep learning systems are used in many areas such as big data analysis, speech recognition, image classification, pedestrian recognition, generic visual recognition and face recognition along with improvements in power of processor and in graphics processors [1].

Deep learning, a class of machine learning uses many nonlinear processing unit layers for feature extraction and conversion. Each subsequent layer calculates the output of the previous layer as input [2].

Algorithms can operate supervised or unsupervised. Deep learning is based on learning with the representation of data. For example; when a representation is represented for an image, properties such as a vector or edge sets of density values per pixel may be considered. Some of these features represent data better, while others may represent worse. Deep learning methods are advantageous by using effective algorithms for hierarchical feature extraction that best represent data rather than manually extracted features [3,4].

The difficulty of generalizing the new examples becomes even more difficult when working with high dimensional data,

and the mechanisms used to achieve generalization in traditional machine learning are not sufficient to learn complex functions in high dimensional areas. Such areas also often require high computational costs. Deep learning is preferred to overcome these and other obstacles [5]. Particularly in the case of large amounts of training data, the deep learning method is an area of considerable work [6-8].

### B. Convolutional Neural Networks (CNN)

In recent years, Convolutional Neural Network (CNN) models, which are deep learning models as an alternative to traditional feature extraction and artificial neural network methods, have started to be developed. CNN-based methods are preferred in order to reduce the effects of different species such as exposure, illumination, facial expressions and better features in image analysis [9].

The CNN has a convolution layer, nonlinearity layer and pooling layers and a flow from simple attributes to complex attributes. Class estimates are obtained as CNN outputs. The filters used in CNN's convolution layers update themselves to minimize the error throughout the training. As the filters are updated, the attributes obtained in the layer outputs also differ. This iterative improvement throughout the training phase provides more distinctive features [1]. The success of this method has been proven, especially in the case of processing large data [6,7,10].

Figure 1 shows a generally accepted CNN model AlexNet architecture.

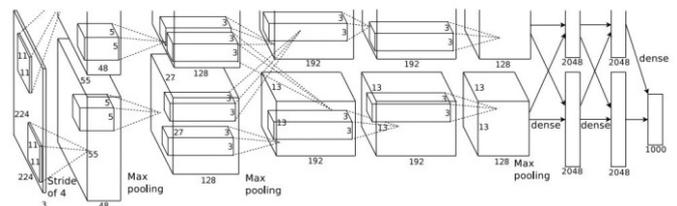


Figure. 1. AlexNet Architecture [11].

## II. PREVIOUS STUDIES

With the increasing amount of data, especially in the image processing area, problems such as image classification and object recognition stand out. Deep learning methods that prove themselves as a method that can cope with large data are frequently used and give the most successful results [4]. However, it is important to increase the amount of data without compromising the integrity of the data format for application areas with a limited set of data.

In their study, Si Liu et al. increased the image size to 1 and 1.2 times in order to increase the sensitivity rate in the cropped human-centered regions and to reduce the excessive compliance, with the filters such as horizontal reflection, 4 variations of each image. However, this study was carried out to include only the image area, and no facial separation was performed [12].

In their work, Vittorio et al. produced a vastly increased image collection to enrich the distinctive features of the painting and tried to eliminate the constraints such as lighting, exposure, facial expression, and low resolution [13]. In their study, Buslaev et al. performed a fast and efficient album creation process by performing multiple image conversion with image enhancement method [14].

Galdran et al. have added color coherence algorithms to the data sets of convolutional neural networks based on convolutional neural network studies for segmentation and classification of skin lesion analysis [15]. Cengil and Çınar performed image classification with CNN model on 8 different images selected from CIFAR-100 data set. In order to be used in the training process of the study, they trained the system they designed for 9 hours with the data increase process [16].

Doğan and Türkoğlu, with deep learning, designed the leaf classification for the model, AlexNet, Vgg16, Vgg19, ResNet50, GoogLeNet algorithms were used to compare the performance. They stated that if the amount of data increases, the performance will increase but the processing time will increase accordingly [17].

In order to increase classification performance in his work, Salman produced artificial data sets by adding different levels of noise through the existing training dataset [18].

In the above mentioned studies, it was emphasized that increasing the data set will increase the educational performance of the system but it is not explained how a performance difference is between the original data set and the increased data set. Increasing the data improperly depending on the type of data used and the application purpose may not yield positive results. The filters used to increase the data set and the proper size of these filters are very important for determining the characteristics.

In this study, on the data set containing few samples; firstly, the classification is made in its original form, then, applying different filters to the data set, the classification has made with the increased new data set. Thus, the performance of data augmentation on the face recognition system has revealed. In addition, it has been tried to determine which filters give more effective results in non-real-time face recognition.

### III. PRESENTED STUDY

#### A. Aim

The purpose of the study is to obtain training data for use on the CNN model, which is designed to obtain a large number of samples from the face images containing a small number of samples in the source dataset without disturbing the integrity of the original image. This will reveal how the non-real-time face recognition system is affected by the classification performance. The process diagram of the proposed study is shown in Figure 2.

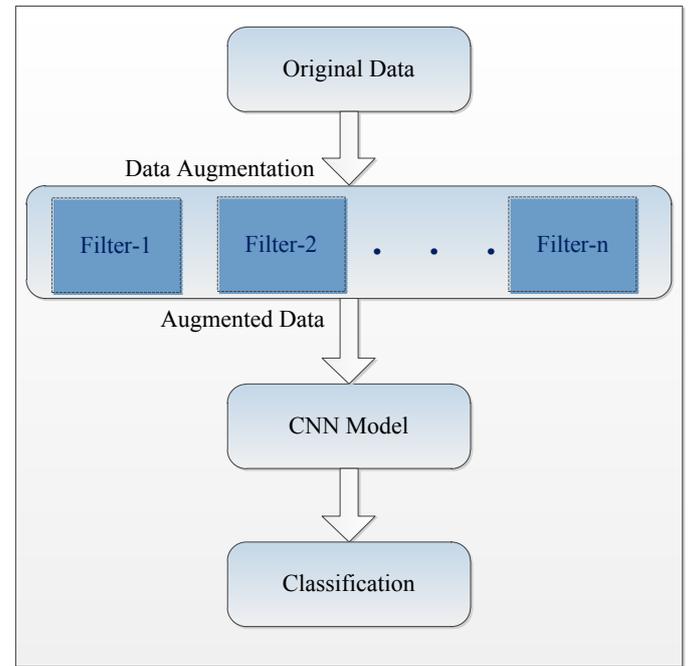


Figure 2. Process Block Diagram

#### B. Method / Architecture

Deep learning architectures include CNN. CNN is a special type of feed forward neural network. The main characteristic of this method is that it can learn features at various levels by providing a more abstract representation of the input data. Thus, it can be applied to the network without attribute extraction. Therefore, a CNN is an end-to-end classifier. Another advantage is the structuring of feature extractors based on data used for training [19].

The processes of the method applied in this study are shown in Figure 3.

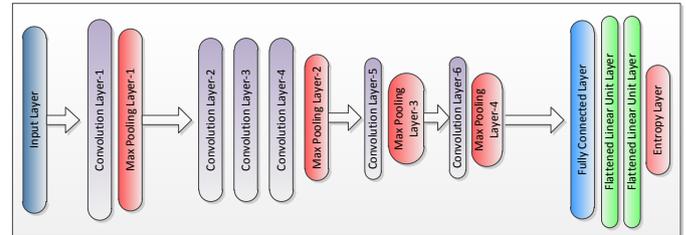


Figure 3. Applied CNN Model

## IV. EXPERIMENTAL SETUP

#### A. Data Sets

150 face images of a small-scale company have been used for the data set. According to the principle of confidentiality of the data, the pictures are not used in their original form. In addition, we have compared the random picture of 150 people with the data set, known as Labeled Faces in the Wild (LFW) [20].

##### 1) Data Preprocessing

All images have taken to 256x256 pixels and the image to be placed on the input layer has obtained in 256x256x3 size.

#### Data Augmentation

The best way to generate a machine learning model, perform better generalization is to train the model using more

data. However, the amount of data we have in practice is limited. One way to overcome this problem is to create artificial data and incorporate it into the training set. This approach is the easiest method for classification. The task of a classifier is to take a complex, high-dimensional input  $x$  and summarize it with a single category  $y$ . That is, the main task of a classifier is to be unchanged against a wide variety of transformations. By converting the  $x$  entries in the training set, new  $(x, y)$  binaries can be created easily [5]. In this way, the variety of data is increased by artificial methods and it affects the learning performance to a great extent. With this process, the pictures are subject to some distortion. On the image, it is attempted to increase the variety of pictures by performing operations such as angular rotation, changing perspective, scrolling and zooming [21].

Many processes, such as rotating or scaling the image, have been shown to be effective in enhancing educational performance. However, it is necessary to avoid making transformations that may perceive the correct class as false. For example, in an optical character recognition system, the letters b and d or numbers 6 and 9 may cause misclassification because they can be likened to each other. It is not appropriate to use data augmentation methods such as horizontal translation or rotation at 180-degree angle for these systems [5].

### B. Implementation Setup

The application has been implemented using the Keras library on the Python platform. To increase the data, Keras's ImageDataGenerator parameters have been used. Using the data enhancement techniques, the number of pictures with separate filters for each face image is gradually increased from 150 to 3300, 6300 and most recently to 9450.

These data sets; angular rotation, scrolling, cropping, zooming, turning filters have been applied and k-nearest neighbor algorithm has chosen to fill the resulting gaps.

TABLE I. AUGMENTED DATA SAMPLES



### C. Calculation of Performance Ratio

The performance of the system has been measured by the accuracy percentage of the training set. The results are presented in Tables I and II. The data quantity and performance ratios obtained from the incremental increase of the original data set are presented in Table II. No data augmentation is applied to compare to 150 images randomly selected from the LFW dataset given in Figure 4. The performance graphs of the original dataset, which are incrementally augmented, are given in Figure 5 and Figure 6.

TABLE II. SUCCESS EVALUATION TABLE

Total Number of Data	Training Time	Performance Rate (%)
150	58sec.	50
3300	20min.	62
6300	44min.	66
9450	1h. 4sec.	76

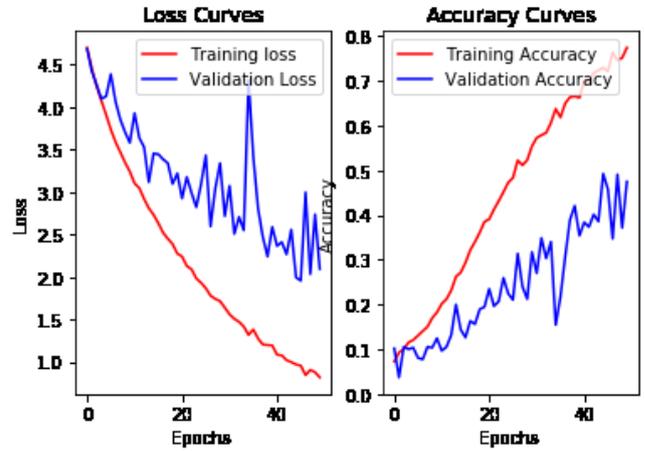


Figure 4. Loss and Accuracy graphs of non-augmented LFW data set.

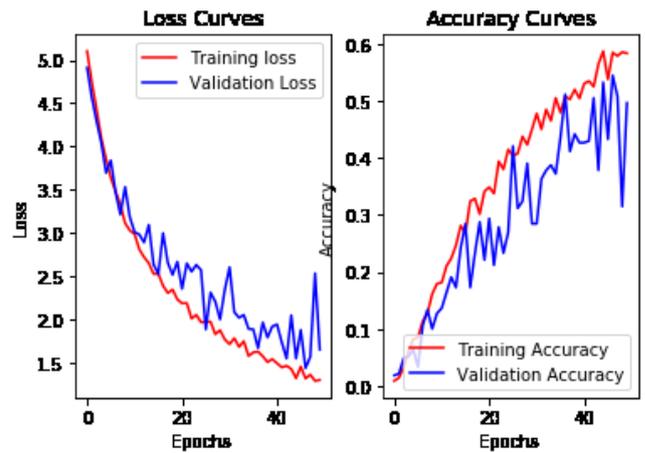


Figure 5. Loss and Accuracy graphs of partial augmented dataset.

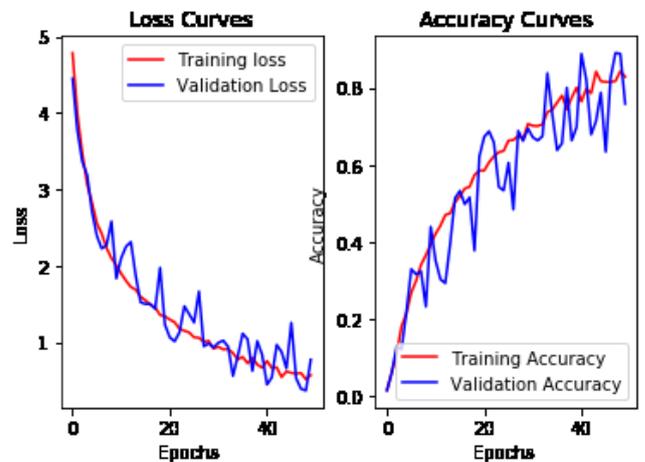


Figure 6. Loss and Accuracy graphs of comprehensive augmented dataset.

## V. CONCLUSION AND EVALUATION

Deep learning systems are used in many areas such as big data analysis, speech recognition, image classification, pedestrian recognition, generic visual recognition and face recognition along with improvements in power of processor and in graphics processors. In particular, in order to extract better attributes in image analysis, traditional feature extraction and an alternative to artificial neural network methods, Convolutional Neural Network (CNN) models, which are deep learning models, have started to be developed.

It is important to increase the amount of data without disrupting the integrity of the data for application areas with limited data set. As the amount of data we have in practice is limited, one way to overcome this problem is to create artificial data and increase this data.

In this study, angular rotation, scrolling, cropping, zooming and turning filters are applied on a small number of data sets and the source data set is increased without disturbing the integrity of the original data. In addition, it has been tried to determine which data augmentation options have more effect on face recognition. Thus, non-real-time face recognition has been performed by training with new augmented dataset of each pictures with many features.

Experimental results show that the performance of the training is significantly increased depending on the amount of data. The filters used in this study, especially angular rotation and brightness filters have more effect on success. Because the LFW dataset contains more than one image for some people, no data augmentation is applied to this data set. This dataset has been used to compare with the augmented original dataset. According to these results, the LFW data set should also be increased for more performance. Although the original data set contains a sample for each person, face recognition is more efficient because only one person is in the pictures. However, the performance achieved with some filters is more effective than others. For example, horizontal flip or mirroring filters may be more suitable instead of vertical flip filter. At the same time, the parameter values used are also very effective. Excessive angular rotation for the data set used did not yield positive results.

## VI. FUTURE STUDIES

In order to further increase training performance, the appropriate filters can be designed to produce more data with reasonable parameter values and new models can be created with AlexNet, ZFNet, GoogLeNet, Microsoft ResNet, R-CNN architectures that can be considered successful in image classification. However, it should be taken into consideration that the training period will increase in proportion to this increase for very high data numbers. In addition, the effects of the filters can be discussed by using different filters than the filters used in this study.

## REFERENCES

- [1] Bilgiç, A. et al., "Face recognition classifier based on dimension reduction in deep learning properties." Signal Processing and Communications Applications Conference (SIU), 2017 25th. IEEE, 2017.
- [2] L. Deng and D. Yu, "Deep Learning: Methods and Applications" Found. Trends® Signal Process., vol. 7, no. 3–4, pp. 197–387, 2014.
- [3] H. A. Song and S.-Y. Lee, "Hierarchical Representation Using NMF" in International Conference on Neural Information Processing, pp. 466–473, 2013.
- [4] Şeker, A. et al., "A Study on Deep Learning Methods and Applications", Gazi Journal of Engineering Sciences, 3.3: 47-64, 2017.
- [5] Ian Goodfellow et al. "Deep Learning", MIT Press, 2016.
- [6] Çalık, N. et al., "Signature recognition application based on deep learning" Signal Processing and Communications Applications Conference (SIU), 2017 25th. IEEE, 2017.
- [7] Ranzato, Y. M. et al., "Sparse Feature Learning for Deep Belief Networks", Proc. Adv. Neural Inf. Process. Syst., vol. 20, pp. 1185-1192, 2007.
- [8] Scherer, D. et al., "Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition", International Conference on Artificial Neural Network, pp. 92-101, 2010.
- [9] Yang, Hu et al., "When Face Recognition Meets with Deep Learning: An Evaluation of Convolutional Neural Networks For Face Recognition", In Proceedings of the IEEE International Conference on Computer Vision Workshops, pp.142-150, 2015.
- [10] Chen, X. W. and Lin X., "Big Data Deep Learning: Challenges and Perspectives", IEEE, vol. 2, pp. 514-525, 2014.
- [11] Krizhevsky, A. et al., "Imagenet classification with deep convolutional neural networks" in Advances in Neural Information Processing Systems 25, pp. 1097–1105. Curran Associates, Inc., 2012.
- [12] Liu, Si et al., "Matching-cnn meets knn: Quasi-parametric human parsing" Proceedings of the IEEE conference on computer vision and pattern recognition, 2015.
- [13] Vittorio, C. et al., "Robust Single-Sample Face Recognition by Sparsity-Driven Sub-Dictionary Learning Using Deep Features", Sensors 19.1:146, 2019.
- [14] Buslaev, A. et al., "Albumentations: fast and flexible image augmentations" arXiv preprint arXiv:1809.06839, 2018.
- [15] Galdran, A. et al., "Data-Driven Color Augmentation Techniques for Deep Skin Image Analysis", arXiv preprint arXiv:1703.03702, 2017.
- [16] Cengil, E. and Çınar, A., "A New Approach for Image Classification: Convolutional Neural Network" European Journal of Technic 6.2, 2016.
- [17] Doğan, F. and Türkoğlu, İ., "Comparison of Leaf Classification Performance of Deep Learning Algorithms", Sakarya University Journal of Computer and Information Sciences 1.1: 10-21, 2018.
- [18] Salman, M., "Integration of Hyperspectral and Lidar Data in Attribute and Decision Levels and Classification with Deep-Curvilinear Neural Networks" Master Thesis, Hacettepe University Institute of Science and Technology, 2018.
- [19] Perlin, H. A. and Lopes, H. S., "Extracting Human Attributes Using A Convolutional Neural Network Approach", Pattern Recognition Letters, Vol. 68, pp. 250-259, 2015.
- [20] Huang, G.B.; Ramesh, M.; Berg, T.; Learned-Miller, E. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments; Technical Report 07-49; University of Massachusetts: Amherst, MA, USA, 2007.
- [21] Şahin, Ö., "TL recognition for visually impaired people on iOS platform", Master Thesis, Selçuk University Institute of Science and Technology, 2017.

# Author Index

- A**  
Abnane, Ibtissam ..... 35  
Achtelik, Markus ..... 341  
Adriani, Matteo ..... 433  
Ahmad, Muhammad Ovais ..... 803  
Aissani, Mohamed ..... 479, 501  
Aleithe, Michael ..... 493  
Alferidah, Saja ..... 293  
Alkhaldi, Nora ..... 293  
Alonazi, Mohammed ..... 627, 633  
Amamra, Abdenour ..... 101, 303, 385, 391  
Amara, Yacine ..... 385, 391  
Amazal, Fatima Azzahra ..... 763  
Amour, Ahmed khierEddine ..... 101  
Andrzejczak, Michał ..... 213  
Artigues, Christian ..... 143  
Atasever, Üyesi Mesut ..... 529
- B**  
Baek, Moo Sang ..... 91  
Baj-Rogowska, Anna ..... 747  
Bakkoury, Zohra ..... 35  
Balduin, Stephan ..... 185  
Baranov, Anton ..... 423  
Barroso, Nuria Rodríguez ..... 255  
Bebeshina-Clairet, Nadia ..... 249  
Belle, Jean-Paul Van ..... 519  
Beloff, Natalia ..... 627, 633  
Benayad, Aissa ..... 391  
Benz, Katharina ..... 907  
Beringer, Grzegorz ..... 273  
Berka, Jakub ..... 881  
Bicevska, Zane ..... 639  
Bicevskis, Janis ..... 639  
Bielecki, Włodzimierz ..... 207  
Biernikowicz, Aneta ..... 659  
Bílek, Petr ..... 891  
Blank-Babazadeh, Marita ..... 185  
Bochenek, Grażyna ..... 591  
Bogach, Natalia ..... 57  
Bouderbal, Imene ..... 303  
Boumaza, Khalid ..... 391  
Bremer, Joerg ..... 193  
Brinkemper, Sjaak ..... 791  
Bryniarska, Anna ..... 19  
Brzoza-Woch, Robert ..... 467  
Buchmann, Erik ..... 449, 555  
Bud'a, Jan ..... 599  
Bumberger, Jan ..... 621  
Burdescu, Dumitru Dan ..... 359
- C**  
Cabodi, Gianpiero ..... 123  
Cámara, Eugenio Martínez ..... 255  
Camurati, Paolo ..... 123  
Carchiolo, Vincenza ..... 605, 645  
Carenzo, Michael ..... 221  
Carôt, Alexander ..... 309  
Catalano, Giovanni ..... 645  
Celejewska-Wójcik, Natalia ..... 591  
Cen, Jiahao ..... 11  
Cen, Ling ..... 7, 15  
Chang, Shuchih Ernest ..... 565  
Cheng, Dongxu ..... 381  
Cherchour, Imed ..... 157  
Chmielarz, Witold ..... 529  
Christensen, Rasmus Engesgaard ..... 277  
Chvatal, Lukas ..... 881  
Costa, Marly ..... 165  
Cybulski, Piotr ..... 471  
Cygańska, Sara ..... 875
- D**  
Daheur, Yasmine ..... 385  
Deng, Bohua ..... 313  
De Sousa, Jr, Rafael T. .... 487  
Dinev, Velizar ..... 43  
Djamaa, Badis ..... 101  
Dolata, Przemysław ..... 29  
Donko, Dženana ..... 755  
Dorodnicov, Sergey ..... 341  
Dörpinghaus, Jens ..... 115, 265  
Drahoš, Peter ..... 925  
Dudycz, Helena ..... 651  
Duic, Neven ..... 613  
Düing, Carsten ..... 115  
Dyk, Rion van ..... 519  
Dziadek, Kamil ..... 67
- E**  
Eilertsen, Alexander Christoffer ..... 277  
Erichsen, Peter Langballe ..... 277  
Esche, Marko ..... 443
- F**  
Fadda, Edoardo ..... 123  
Fernández, José Alberto ..... 255  
Fidanova, Stefka ..... 177  
Fietkau, Julian ..... 493  
Filho, Cicero Costa ..... 165  
Filho, Francisco L. de Caldas ..... 487  
Franczyk, Bogdan ..... 459, 493

<b>G</b> abryelczyk, Renata .....	659	Kosmol, Linda .....	685
Gaida, Paulina .....	841	Kostolani, Michal .....	911
Ganzha, Maria .....	177	Kottke, Mario .....	851
Geigis, Max .....	333	Kowalska, Aleksandra .....	427
Gerasimovich, Aleksandr .....	327	Kowalski, Tomasz .....	427
Głowacki, Mirosław .....	591	Kozák, Štefan .....	911
Gong, Weiyong .....	313	Kramarczyk, Michał .....	349
Grad, Łukasz .....	3	Krawiec, Łukasz .....	651
Grimmer, Martin .....	459	Kreußel, Dennis .....	459
Grzegorowski, Marek .....	3	Książopolski, Bogdan .....	231
<b>H</b> aar, Christoph .....	449	Kučera, Erik .....	317, 915, 925
Hadhbi, Youssouf .....	127	Kulpa, Artur .....	705
Haeri, Seyed Hossein .....	399	Kuo, Tzu-Yin .....	565
Haffner, Oto .....	317, 915, 925	Kutlugün, Mehmet Ali .....	929
Han, Hyo Chang .....	497	<b>L</b> afourcade, Mathieu .....	249
Hanslo, Ridewaan .....	813	Laszczyk, Maciej .....	47, 67
Hebrard, Emmanuel .....	143	Lehnhoff, Sebastian .....	185, 193
Herrera, Francisco .....	255	Leskovský, Roman .....	317, 915
Hoene, Christian .....	309	Levina, Alla .....	227
Hoffmann, Jörn .....	459	Leyh, Christian .....	685
Hubl, Marvin .....	493, 505	Lezhenin, Iurii .....	57
<b>I</b> dri, Ali .....	35, 763	Ligeza, Antoni .....	733
Idrissi, Touria El .....	35	Lin, Jung-Hsin .....	135
Ilyés, Enikő .....	823	Lipcák, Peter .....	771
<b>J</b> abłoński, Janusz .....	547	Lipka, Richard .....	781
Jabłoński, Mateusz .....	273	Liu, Cenru .....	11
Jankowski, Jarosław .....	663	Longheu, Alessandro .....	605
Janusz, Andrzej .....	3	Łuczak, Piotr .....	427
Januszewski, Piotr .....	273	Łukasiewicz, Katarzyna .....	875
JiSeon, Yun .....	743	Luque, Gabriel .....	177
Juhár, Ján .....	411	Lyakhovets, Dmitriy .....	423
Jurková, Tereza .....	599	<b>M</b> acák, Martin .....	771
<b>K</b> abardov, Muaed .....	327	Macik, Miroslav .....	891
Kania, Aleksander .....	591	Mahmood, Hasan .....	309
Karabegovic, Almir .....	613	Maiza, Mohamed .....	157
Karakaya, Mehmet Ali .....	929	Malgeri, Michele .....	645
Karczmarczyk, Artur .....	663	Manerba, Daniele .....	123
Karolyi, Matěj .....	599	Marcinkiewicz, Michał .....	61
Kasal, Pavel .....	107	Marnier, Kristina .....	831
Kasprzak, Włodzimierz .....	363	Matišák, Jakub .....	317, 921
Keir, Paul .....	399	Mazalová, Monika .....	599
Kerivin, Hervé .....	127	Meigen, Christof .....	621
Kęsik, Karolina .....	287	Meixner, Gerrit .....	901
Kfouri, Guilherme Oliveira .....	487	Mendelson, Haim .....	535
Khmeliuk, Volodymyr .....	539	Mendonça, Fabio L. L. .....	487
Kim, Young Myung .....	725	Mercorelli, Paolo .....	907
Klein, Achim .....	43, 569	Merkt, Oana .....	693
Kluza, Krzysztof .....	733	Měšťák, Jan .....	107
Komenda, Martin .....	599	Michalec, Tomasz .....	467
Korczak, Jerzy .....	675	Mikovec, Zdenek .....	881, 891
Korzhik, Valery .....	327	Miler, Jakub .....	841
		Miller, Gloria .....	717
		Mireles, Gabriel Alberto García .....	861

Miśkiewicz, Marek	231	Respondek, Jerzy	87
Mnkandla, Ernest	813	Riekert, Martin	43
Morales-Luna, Guillermo	327	Riel, Andreas	851
Moya, Antonio R.	255	Robak, Marcin	555
Mrukwa, Grzegorz	61	Robak, Silva	547
Mrzygłód, Barbara	591	Roeva, Olympia	177
Mucherino, Antonio	135	Röhling, Martin Max	459
Murín, Justín	911	Romanowski, Andrzej	427
Muszyńska, Karolina	705	Romero, Elena	255
Myszkowski, Paweł B.	47, 67	Rose, Dennis Højbjerg	277
<b>N</b>		Rosinová, Danica	317
Nakasho, Kazuhisa	77	Rossi, Bruno	771
Naldi, Maurizio	433	Rouigueb, Abdenebi	157
Nastafek, Paweł	591	Ruta, Dymitr	7, 15
Nath, Rudra Pratap Deb	277	Ryaskin, Gleb	227
Nguyen, Diep	901	<b>S</b>	
Ng, Yen Ying	871	Safwat, Sherine	373
Nieße, Astrid	185	Saichi, Lamia	385
Nikiforova, Anastasija	639	Salem, Mohammed Abdel-Megeed	373
Novosel, Tomislav	613	Salwiczek, Felix	443
Nowakowski, Grzegorz	539	Sankowski, Dominik	427
<b>O</b>		Savu, Beniamin	355
Oditis, Ivo	639	Šcavnický, Jakub	599
Oest, Frauke	185	Schneider, Stefan	333
Ohzeki, Kazuo	333	Senouci, Mustapha Reda	479, 501
Olszewska, Joanna Isabelle	81	Shabanov, Boris	423
Opaliński, Andrzej	591	Shen, Yuanyuan	535
Overbeek, Sietse	791	Sielski, Dawid	427
<b>P</b>		Silva, Daniel Alves da	487
Pąkowski, Marek	207	Silva, Sergio	165
Paszun, Tymoteusz	733	Şirin, Yahya	929
Pellegrino, Carlo	645	Skowron, Philipp	493
Pinheiro, Alexandre	487	Sładek, Krzysztof	591
Platania, Giulio	645	Sobecki, Andrzej	273
Pohl, Daniel	341	Song, Ha Yoon	91, 497, 725, 743
Polak, Monika	221	Sroka, Wiktor	675
Poław, Dawid	287	Stanescu, Liana	355
Pórolniczak, Edward	349	Stark, Erich	317, 925
Pondel, Maciej	675	Starostin, Vladimir	327
Ponjavic, Mirza	613	Šťastná, Jana	237
Poth, Alexander	851	Stefan, Andreas	265
Potuzak, Tomas	781	Stefańczyk, Maciej	363
Praciano, Bruno J. G.	487	Štěpánek, Lubomír	107
Pyshkin, Evgeny	57	Sung, Minsuk	91
<b>Q</b>		Swacha, Jakub	705
Quiliot, Alain	143	Szydło, Tomasz	467
Quinn, M.	81	Szymański, Julian	273
<b>R</b>		<b>T</b>	
Rábek, Matej	921	Tadei, Roberto	123
Rapoport, Michael	147	Tamir, Tami	147
Raulamo-Jurvanen, Paivi	803	Tashkandi, Alalaa	711
Rech, Claus	907	Telegin, Pavel	423
Regulski, Krzysztof	591	Telenyk, Sergii	539
Reiner, Jacek	29	Theobald, Sven	831
Reitano, Giuseppa	605	Tkout, Abderahmane	157

Toro, Federico Grasso .....	443	William, Youssef .....	373
Torres, José Alberto Sousa .....	487	Winnicka, Alicja .....	287
Toussaint, H�el�ene .....	143	Wi�niewski, Piotr .....	733
Trapani, Natalia .....	645	Wojczuk, Maksymilian .....	467
Trujillo, Miguel Eh�catl Morales .....	861		
Tseng, Min-Han .....	565	<b>X</b> u, Xuebin .....	313, 381
Tudoroiu, Nicolae .....	359		
Tudoroiu, Roxana-Elena .....	359	<b>Y</b> achir, Ali .....	101
<b>V</b> ahed, Anwar .....	813	Yakovlev, Victor .....	327
Vogel, Mandy .....	621	Yefremov, Kostiantyn .....	539
Vu, Quang Hieu .....	7, 15		
<b>W</b> achter, Philipp .....	569	<b>Z</b> afer, Mostefa .....	479, 501
Wagler, Annegret .....	127	Zagarella, Luca .....	605
Wagner, Stefan .....	831	Zaheeruddin, Mohammed .....	359
Wallrafen, Susanne .....	493	�akov, Katarna .....	921
Wtrbski, Jaroslaw .....	663	Zborowski, Marek .....	529
Wawryk, Maciej .....	871	Zerrouki, Ali .....	101
Weichbroth, Pawe .....	747	Zhang, Xinman .....	313, 381
Weil, Vera .....	115	Zhuvikin, Aleksey .....	327
Westwanska, Weronika .....	87	Zielonka, ukasz .....	171
White, Martin .....	627, 633	Ziomba, Ewa .....	579
Widmer, Tobias .....	569	Zikratov, Igor .....	227
Wilkowski, Artur .....	363	Zimpel, Tobias .....	505
		�uni, Emir .....	755