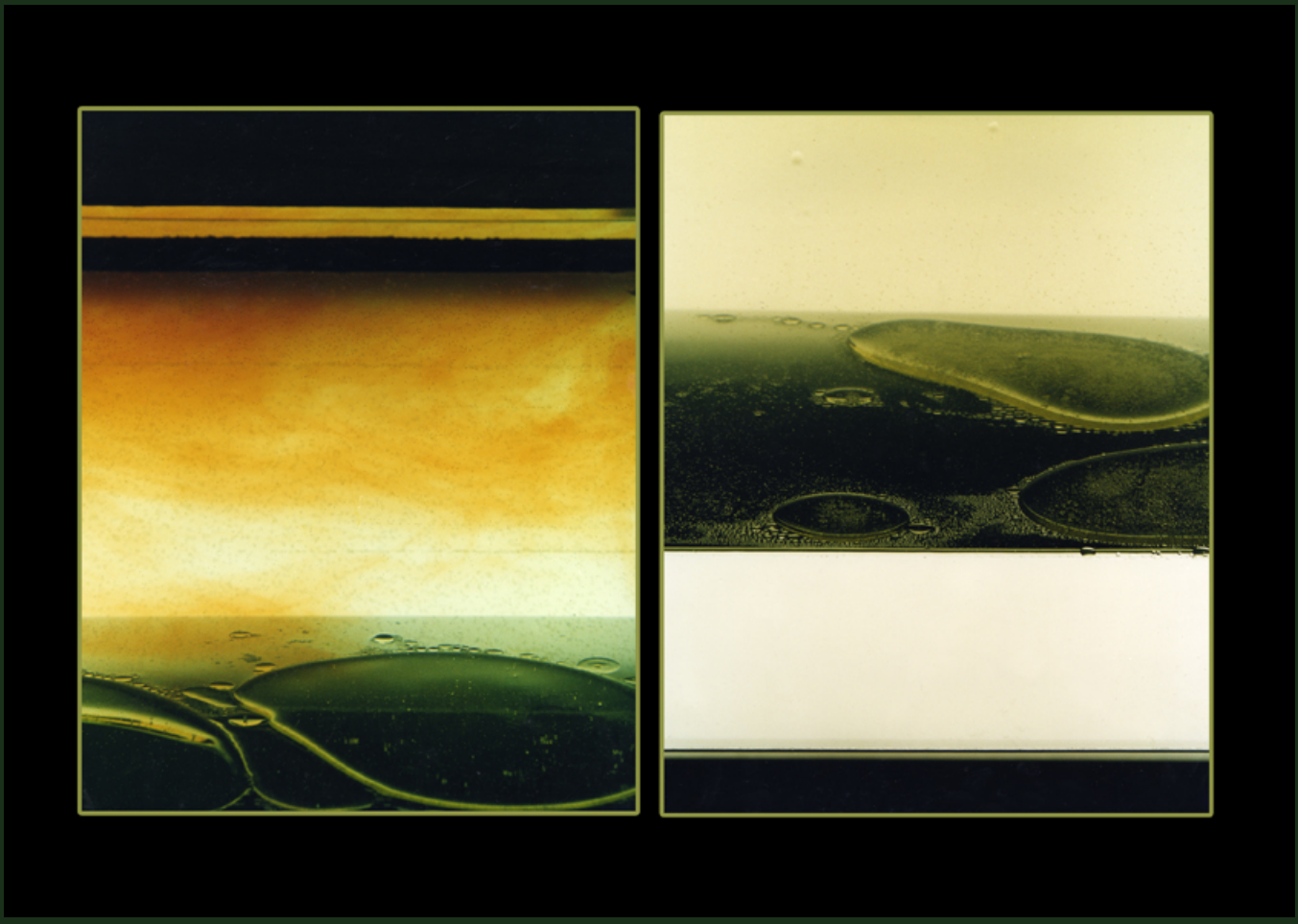
Annals of Computer Science and Information Systems
Volume 20

Communication Papers of the 2019 Federated Conference on Computer Science and Information Systems

September 1–4, 2019. Leipzig, Germany



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

PTI

Annals of Computer Science and Information Systems, Volume 20

Series editors:
Maria Ganzha (Editor-in-Chief),
Systems Research Institute Polish Academy of Sciences and Warsaw University of Technology, Poland
Leszek Maciaszek,
Wrocław Universty 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,
Nałęcz 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

TEXnical editor: Aleksander Denisiuk,
University of Warmia and Mazury in Olsztyn, Poland

Communication Papers of the 2019 Federated Conference on Computer Science and Information Systems

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

2019, Warszawa, Polskie Towarzystwo Informatyczne

Annals of Computer Science and Information Systems, Volume 20
Communication Papers of the 2019 Federated Conference on Computer Science and Information Systems

USB: ISBN 978-83-955416-4-3
WEB: ISBN 978-83-955416-3-6

ISSN 2300-5963
DOI 10.15439/978-83-955416-3-6

© 2018, Polskie Towarzystwo Informatyczne
Ul. Solec 38/103
00-394 Warsaw
Poland

Contact: secretariat@fedcsis.org
http://annals-csis.org/

Cover photo:
Iwona Burdecka,
Elbląg, Poland

Also in this series:

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 18: Proceedings Papers of the 2019 Federated Conference on Computer Science and Information Systems, **ISBN Web 978-83-952357-8-8, ISBN USB 978-83-952357-9-5, ISBN ART 978-83-955416-0-5**
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 Papers 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**

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

The communication papers were introduced in 2017 as a separate category of contributions. They report on research topics worthy of immediate communication. They may be used to mark a hot new research territory or to describe work in progress in order to quickly present it to scientific community. They may also contain additional information omitted from the earlier papers or may present software tools and products in a research state.

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 were 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 (7th 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)

Each paper, found in this volume, was refereed by at least two referees.

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 298 regular submissions.

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. Organizing 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, Bułgaria.

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, *Wrocław University of Economics, Wrocław, Poland and Macquarie University, Sydney, Australia*
Marcin Paprzycki, *Systems Research Institute Polish Academy of Sciences, Warsaw Poland and Management Academy, Warsaw, Poland*

Communication Papers of the 2019 Federated Conference on Computer Science and Information Systems (FedCSIS)

September 1–4, 2019. Leipzig, Germany

TABLE OF CONTENTS

ARTIFICIAL INTELLIGENCE AND APPLICATIONS

14TH INTERNATIONAL SYMPOSIUM ADVANCES IN ARTIFICIAL INTELLIGENCE AND APPLICATIONS

Call For Papers **1**

Comparative Analysis of Data Mining Algorithms Applied to the Context of School Dropout **3**
Nathanael Oliveira Vasconcelos, Methanias Colaço Júnior, Thiago S. Almeida, Victor Matheus da Silva

12TH INTERNATIONAL WORKSHOP ON COMPUTATIONAL OPTIMIZATION

Call For Papers **11**

Selecting representatives **13**
László Aszalós, Dávid Nagy

Correlation Clustering: Let All The Flowers Bloom! **21**
László Aszalós, Mária Bakó

Exact and approximation algorithms for joint routing and flow rate optimization **29**
Konstanty Junosza-Szaniawski, Dariusz Nogalski

COMPUTER SCIENCE & SYSTEMS

Call For Papers **37**

12TH WORKSHOP ON COMPUTER ASPECTS OF NUMERICAL ALGORITHMS

Call For Papers **39**

Symbolic description of the polynomial roots and their numerical implementation - better than in Mathematica software? **41**
Edyta Hetmaniok, Damian Słota, Mariusz Pleszczyński, Roman Wituła, Michał Różański, Marcin Szczygieł

12TH INTERNATIONAL SYMPOSIUM ON MULTIMEDIA APPLICATIONS AND PROCESSING

Call For Papers **47**

The Use of Modern Multimedia Trends for Popularization of Mechatronics by Presenting Its Present and Future Application **49**
Erik Kučera, Oto Haffner, Roman Leskovský, Jakub Matišák, Peter Drahoš

NETWORK SYSTEMS AND APPLICATIONS

ADVANCES IN NETWORK SYSTEMS AND APPLICATIONS

Call For Papers **55**

A Framework for Network Intrusion Detection using Network Programmability and Data Stream Clustering Machine Learning Algorithms **57**
Admilson de Ribamar Lima Ribeiro, Edward David Moreno Ordonez, Anderson Clayton Alves Nascimento

3RD WORKSHOP ON INTERNET OF THINGS - ENABLERS, CHALLENGES AND APPLICATIONS

Call For Papers **65**

Solutions for Planning Smart Hybrid Public Transportation System - Poznan Agglomeration as a Case Study of Satellite Towns' Connections **67**
Marzena Banach, Katarzyna Kubiak, Rafał Długosz

Positioning Improving of RSU Devices Used in V2I Communication in Intelligent Transportation System **73**
Rafał Długosz, Marzena Banach, Katarzyna Kubiak

Design of a Distributed HIDS for IoT Backbone Components **81**
Guilherme de O. Kfouri, Daniel G. V. Gonçalves, Bruno V. Dutra, João F. de Alencastro, Francisco L. de Caldas Filho, Lucas M. C. e Martins, Bruno J. G. Praciano, Robson de O. Albuquerque, Rafael T. de Sousa Jr

INFORMATION SYSTEMS AND TECHNOLOGY

Call For Papers **89**

17TH CONFERENCE ON ADVANCED INFORMATION TECHNOLOGIES FOR MANAGEMENT

Call For Papers **91**

Efficient Production Monitoring on the Basis of Domain Ontologies by Utilizing IoT **93**
Maha Riad, Amal Elgammal, Doaa Elzanfaly

A mixed-methods measurement and evaluation methodology for mobile application usability studies **101**
Paweł Weichbroth

14TH CONFERENCE ON INFORMATION SYSTEMS MANAGEMENT

Call For Papers **107**

Improved Virus Optimization algorithm for two-objective tasks Scheduling in Cloud Environment **109**
Kadda Beghdad Bey, Sofiane Bouznad, Farid Benhammadi, Hassina Nacer

Building a Clinical Case Reports Database to Facilitate Case-Based Learning and Improve Critical Thinking of Medical Students **119**
Jaroslav Majerník, Peter Jarčuška, Jozef Gonsorčík

25TH CONFERENCE ON KNOWLEDGE ACQUISITION AND MANAGEMENT

Call For Papers **123**

Innovative Modular Approach based on Vehicle Routing Problem and Ant Colony Optimization for Order Splitting in Real Warehouses **125**
Emir Žunić, Sead Delalić, Zlatan Tucaković, Kerim Hodžić, Admir Beširević

SOFTWARE AND SYSTEM ENGINEERING

JOINT 39TH IEEE SOFTWARE ENGINEERING WORKSHOP (SEW-39) AND 6TH INTERNATIONAL WORKSHOP ON CYBER-PHYSICAL SYSTEMS (IWCPS-6)

Call For Papers **131**

A roadmap for a comparison framework for an adaptable software process improvement framework in small software companies **133**
Micheal Tuape, Yirsaw Ayalew

Author Index **143**

14th International Symposium Advances in Artificial Intelligence and Applications

AAIA'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
- **Błaszczyński, Jerzy,** Poznań University of Technology, Poland
- **Calpe, Javier**
- **Camastra, 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.,** Wroclaw University of Technology, Poland
- **Nowostawski, Mariusz,** Norwegian University of Technology and Science (NTNU), Norway
- **Ohsawa, Yukio,** University of Tokyo, Japan
- **Olech, Łukasz,** Wroclaw 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,** Wroclaw 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,** Wroclaw 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**
- **Ślużek, 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

Communication Papers of the Federated Conference on Computer Science and Information Systems pp. 3–10
DOI: 10.15439/2019F265
ISSN 2300-5963 ACSIS, Vol. 20

Comparative Analysis of Data Mining Algorithms Applied to the Context of School Dropout

Nathanael Oliveira Vasconcelos*, Methanias Colaço Júnior†, Thiago S. Almeida‡ and Victor Matheus da Silva §
*‡§ Federal University of Sergipe, Brazil
*†‡ Postgraduate Program in Computer Science - PROCC
*Email: nathanvasc@gmail.com
†Competitive Intelligence Research and Practice Group – NUPIC
†Email: mjrse@hotmail.com
‡ Email: thiago7a@gmail.com
§Email: victormsti@gmail.com

***Abstract*—Students' dropout is certainly one of the major problems that afflict educational institutions, the losses caused by the student's abandonment are social, academic and economic waste. The quest for its causes has been subject of work and educational research around the world. Several organizations seek strategic decisions to control the dropout rate. This work's goal is to evaluate the effectiveness of the most used data mining algorithms in the education area. An "in vivo" controlled experiment was planned and performed to compare the efficacy selected classifiers. The Random Forest and SVM algorithms have stood out in this context, having, statistically similar accuracy (80.36%, 81.18%), precision (80.79%, 80.25%), recall (76.50%, 77.51%) and f-measure (78.86%, 78.81%) averages. The results showed evidence of significant differences between the algorithms, and also showed that, although the SVM had the best metric of accuracy and recall, it results were statistically similar with Random Forest results.**

I. Introduction

According to the 2015 Higher Education Census in Brazil, 11% of the students entering the undergraduate program in 2010 dropped out in the first year. By 2014, almost half (49%) of students had left the courses they had opted for in 2010 [8]. Thus, dropout is certainly one of the major problems that afflict educational institutions in general. The search for its causes has been the object of many studies and educational researches [32] [22] [21][14].

Several brazilian governmental organizations such as the REUNI Program - Reestruturação e Expansão das Universidades Federais (Restructuring and Expansion of Federal Universities) - and the TAM - Termo de Acordo de Metas e Compromissos (Goals and Commitments Agreement Term) - seek for strategic decisions that try to control the dropout rate [32]. The losses caused by the student's abandonment are social, academic and economic waste. In the public sector, resources are invested without due return, in parallel, in the private sector, evasion rates represent a significant loss of revenue. Consequently, there is the need to increase the understanding of the problem and its causes, by adopting more effective measures to identify and understand the main factors that can contribute to student failure.

A very promising information gatthering alternative is the use of "knowledge discovery in databases" and the use of "data mining techniques in education", also called Educational Data Mining (EDM) [22].

EDM is defined by "The Educational Data Mining" community website[1] as an emerging discipline, concerned with the development of methods to explore the unique types of data that come from educational environment and use those methods to better understand the students and the characteristics of their learning.

Similarly, it is possible to mine data from students in order to identify relationships among the various factors that lead them to abandon the course. However, predicting school dropout is a multifactorial problem that includes several variables such as family, social, economic, and personal factors [27].

This work's goal is to evaluate the effectiveness of the most used data mining algorithms in the educational area for the prediction of students on the verge of academic dropout in the context of a public higher education institution. The results show evidence that the Random Forest and SVM algorithms have stood out in this context, having, statistically similar accuracy (80.36%, 81.18%), precision (80.79%, 80.25%), recall (76.50%, 77.51%) and f-measure (78.86%, 78.81%) averages.

This paper is organized as follows. Section II presents the related works. Section III presents the methodology used in this research. Section IV presents the required theoretical concepts to understand the research. Section V presents experiment definition and planning, followed by the presentation of its operation in Section VI. Section VII presents the results, and final considerations can be found in Section VIII.

II. Related Works

Recently, a reasonable number of researches have been conducted to apply data mining techniques in the education area, in order to classify and predict student performance in various education institutes [10] [22] [21] [17] [20]. The use of such techniques in education is promising due to the ammount of oportunities in this area [1].

[1] www.educationaldatamining.org

©2019, PTI

Iam-on and Boongoen [14] present as a study case the Mae Fah Luang University, in Thailand, by using EDM models and proposing a new data transformation approach to improve the accuracy of conventional classifiers aiming at the disseminating of interesting patterns with a higher accuracy. Their works contributed to predict students' performance and possible dropouts, based on their pre-university characteristics, admission details, and initial academic performance at the university. The limitation to their model is the complexity and the required time, so it may not work well with larger datasets.

Dekker et al. [10] were able to identify, in the first school year, the students who presented the highest risk of dropout. The study considered several students' data and obtained accuracy between 75% and 80% using a tree decision classifier.

Márquez-Vera et al. [22] propose the application of data mining techniques to predict school failure and dropout in a case study with data from 670 high school students in Zacatecas, Mexico. The acuracies obtained ranged from 75% to 98%, considering ten classifiers. In this study, students' scores were used with greater emphasis, in relation to other attributes. The authors conclude that classification algorithms can be used successfully in order to predict students' development. It is worth noting that despite having a high accuracy in some trials, the context applied is different from the one proposed in this work.

Some EDM studies in the scope of the Brazilian school dropout are highlighted, and we will discuss them in the next paragraphs.

Manhães et al. [21] compared 6 classifying algorithms and found problems with students who can not complete their undergraduate courses. The data sample is composed of 7304 students from higher education course at UFRJ - Universidade Federal do Rio de Janeiro. The data are classified into three classes: students who completed the course obtaining the diploma, students who could not complete the course, and students who had active enrollment after the average deadline for the conclusion of the undergraduate course at the federal institution. The study obtained an accuracy precision of around 80%.

Pascal et al. [17] addresses the dropout rate in a graduation course in a public higher education institution, considering only the Business Management and Zootechny courses. The research uses several machine-learning methods for prediction, and its tests have reached an accuracy higher than 70 % in students dropout prediction.

Machado et al. [20] presented in the article "Bibliometric study in data mining and school dropout", a bibliometric survey of articles published between 2005 and March 2015 that address data mining and school dropout issues. The survey was conducted by using the *Scopus*, *Web Of Science*, *Science Direct* and *Scielo* databases.

The search had as result 16 articles from *Scopus* database, 6 from *Web of Science*, 3 from *Sience Direct* and none from *Scielo* databases. Therefore, 24 scientific articles were part of this study scope.

Machado et al. [20] also listed nine data mining methods found in their research, among them *Decision Trees*, *Neural Network*, *Rule Induction* and *Support Vector Machine*.

In Table I, it is possible to observe a list with the previously cited related works, their classifiers and the obtained accuracies. It is important to note that the papers use different databases for their studies, making it impossible to compare the results directly. However, some of these studies have also used data from a public higher education institution, and along with our article, may help in a secondary general analysis.

It is worth mentioning that there were no studies that performed a comparative analysis of algorithms applied to the school dropout context considering an experimental approach with statistical validation of the significance of data, as proposed in this paper. A robust knowledge base can only be generated with the replications of real controlled experiments that statistically validates their work, which can serve as input for real data meta-analyzes.

III. Methodology

The methodology used in this work, in terms of classification, consists of an exploratory research [29], as a literature review was conducted with systematic approaches, in which the attributes used to generate the files used by the mining tool were defined. The selection of attributes was performed after analysis of the related works, like work [22], as well as by considering the attributes available in the database.

In addition, we used the attribute selection algorithm Random Forest, which evaluates the predictive value of each attribute individually, generating a ranking in which those attributes that have more relation with the class and less correlation with the other attributes receive higher scores [31].

After defining the attributes, mining models were generated, in order to make a comparison between the used algorithms. The main classifiers found in the EDM works [25][22] are: Decision Tree [5], Naive Bayes [28], Nearest Neighbor (KNN) [6], Neural Networks (MLP) [24], Support Vector Machines (SVM) [7] and Ensemble Methods (Random Forest)[4]. According to Xindong et al. [34] these algorithms are among the most used ones in data mining.

Finally, to achieve this research main goal and subsequent data collection, a controlled "in vivo" experiment was proposed and carried out, which involved the database of a public higher education institution. According to Wohlin et al.[33], experimentation is not a simple task, as it involves preparing, conducting and analyzing data correctly. The authors highlight the control of subjects, objects and instrumentation as one of the main advantages of the experimentation, which makes it possible to draw more general conclusions on the investigated subject.

Other advantages include the ability to perform statistical analyzes by using hypothesis testing methods and opportunities for replication. Juristo et al. [16] also state that scientific research can not be based on commercial opinions or interests. Scientific investigations are represented by studies based on observation and/or experimentation with the real world and

TABLE I
LIST OF THE MAIN RELATED WORKS, CLASSIFIERS USED AND ACCURACIES

Author	Classifiers	Accuracies
[10]	OneR, CART, Decision Tree, Naive Bayes, Net Logit, JRip, Random Forest	75% to 80%
[22]	JRip, NNge, OneR, Prism, Ridor, Decision Tree, SimpleCart, ADTree, RandomTree, REPTree	75% to 99%
[21]	Decision Tree, Support Vector machine (SVM), AdaBoost, Naïve Bayes, SimpleCart, MLP	Average = 80%
[17]	Decision Tree, KNN, CART, Naive Bayes, MLP	Average = 74%

their measurable behaviors, as in this research. These aspects should also be taken into account in the construction and evaluation of algorithms and software.

In the execution of the experiment, with the definition of the algorithms and attributes, the Python language and its libraries were used, with this, knowledge models were generated in order to perform algorithm tests and compare the effectiveness. In this context, this work was also classified as laboratory and experimental, due to the planning and execution of a controlled experiment.

To assist the calculations and to verify if there were significant differences in the algorithms efficiency, the Statistical Package for Social Science - SPSS [15] tool was used for data analysis, applying basic and advanced statistical techniques. The SPSS is a statistical software used internationally for many decades, since its versions for large computers [23].

In summary, the experiment can be divided into four main stages: planning with a selection of attributes; data cleaning operation, dataset generation and data collection; comparison of algorithms; and finally the results analysis. The experiment in question is detailed in Sections V and VI.

IV. CONCEPTUAL BASE

In this section, some concepts that are necessary for the understanding of this work are presented.

A. Data Classification

Classification is the process of associating specific objects (instances) into a set of categories (classes or concepts), based on their object properties. Classification is a procedure where individual items are placed in groups based on information derived from characteristics inherent in the items and based on a training set previously labeled [11]. The algorithms used in this research are cited in Section III.

B. Matrix of Confusion

Among the various ways of evaluating a classifier's predictive ability to determine the class of multiple records, the confusion matrix is the simplest of these forms [12].

For n classes, the confusion matrix is a dimension table $n \times n$. For each possible classification, there is a corresponding row and column, it means, the values of the classifications will be distributed in the matrix according to the results, thus generating the confusion matrix for the prepared classifications [3]. The rows correspond to the correct classifications, and the columns represent the classifications performed by the classifier [13] .

When there are only two classes, one is considered *positive* (in the context of this work, "Evaded") and the other as *negative* [13]. Thus, we can have four possible outcomes:

- *True Positive* (TP): a positive class instance is correctly classified as positive;
- *Negative* (FN): a positive class instance is incorrectly classified as negative;
- *True Negative* (TN): a negative class instance can be correctly denoted as negative;
- *Positive* (FP): a negative class instance is incorrectly classified as positive.

C. Quality Metrics

In this work, the accuracy, recall, precision and F-measure metrics were used.

1) Accuracy: It is the percentage of instances sorted correctly.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

2) Recall: It is the percentage of instances that were correctly classified as positive.

$$recall = \frac{TP}{TP + FN} \tag{2}$$

3) Precision: It is the percentage of instances rated positive that are really positive.

$$precision = \frac{TP}{TP + FP} \tag{3}$$

4) F-Measure : Also known as harmonic measure, because it combines precision and recall, evenly weighting.

$$F - measure = \frac{2.precision.recall}{precise + recall} \tag{4}$$

V. EXPERIMENT PLANNING

In this and the next two sections, this work is presented as an experimental process. It follows the [33] guidelines. This section will focus on goal setting and experiment planning. Figure 1 illustrates the steps of the work, this section will focus on step 1, that is, goal setting and experiment planning.

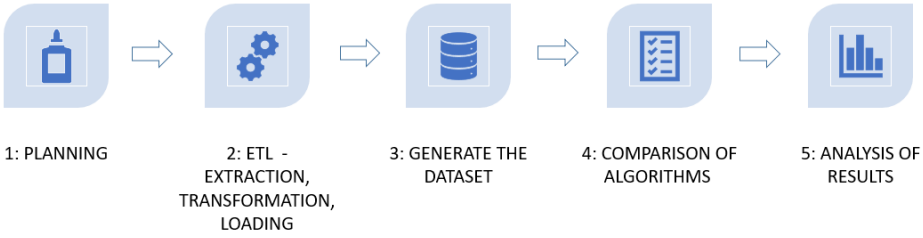


Fig. 1. Stages of the work.

A. Goal Setting

The goal of this study is to evaluate the main classifiers found in EDM, identifying the best algorithm in terms of effectiveness, focusing on school dropout, within a higher education institution. The major models will be used to form a predictive metamodel that will help with dropout management.

The experiment will target undergraduate students from a higher education institution. The goal was formalized by using the GQM model proposed by [3]: **Analyze**, through a controlled experiment, the main algorithms of data mining applied to the context of education, with **focus** on school dropout, **for the purpose** of confirming and/or identifying the best algorithm in terms of efficacy, **with respect to** accuracy, recall, precision and f-measure, **from the point of view** of researchers and data analytics professionals, **in the context** of data on the dropout rate of a public institution of higher education.

1) Context Selection: The experiment was "in vivo" and considered the data of students of all undergraduate courses, from a public higher education institution, with admittance between 2003, year in which the first undergraduate courses began, and 2017. The year/period of the institution during the period of preparation of this article was 2018/1, which, for this reason, was not part of this study. Data selection took into account personal, academic, and social attributes.

2) Hypothesis Formulation: To guide the study, the following research question was elaborated, whose answer aims to fulfill the objective of the work. In the context of the School Dropout Rates in a higher education institution, among the algorithms selected in the EDM area, which is the best in terms of Efficacy?

To evaluate this question, we used four metrics: Accuracy (1), Recall (2), Precision (3) and F-Measure (4). Thus, with the objectives and metrics defined, the following hypothesis will be considered (for each metric):

- H_0: The algorithms$_{(1,2..n)}$ have the same averages for the metric.
  $\mu 1$(metric) = $\mu 2$ (metric) ... = μ n (metric);
- H_1: The algorithms$_{(1,2..n)}$ have different averages for the metric.
  $\mu 1$ (metric) $\neq$ $\mu 2$ (metric) ... $\neq$ μ n (metric);

3) Selecting Participants: All undergraduate students of a higher education institution in all the complete academic periods were considered. This selected institution is public and has several courses of different levels. The database analyzed was the SIGAA - Sistema Integrado de Gestão de Atividades Acadêmicas (Integrated Academic Activities Management System), which stores the entire academic life of the institution's students. The institution provided the database for the experiment in question.

4) Independent variables: For the classification task, we considered 17 attributes, from the base described in subsection V-A3, which are presented in the Table II. The used algorithms are: Decision Tree, Naive Bayes, k-Nearest Neighbor (KNN), Neural Networks (MLP), Support Vector Machine (SVM) and Ensemble Methods (Random Forest), with the parameters presented in the Table III.

TABLE II
ATTRIBUTES CONSIDERED FOR ANALYSIS

Attribute	Description
sexo	Student gender
idade	Student age at the beginning of the course
inst_seg_grau	High school institution type
raca	Student Ethnicity
est_civil	Marital status
qtd_tranc	Number of stopouts in the course
reab_matricula	Indicates whether the student has re-enrolled in the course
qtd_ap_med_p	Average number of courses approved per period
qtd_ap_1p	Number of courses approved in the first period
qtd_rep_med_p	Average number of failed subjects per period
qtd_rep_1p	Number of failed subjects in the first period
qtd_per_cur	Number of periods attended by the student
cra	Academic performance coefficient
perc_aprov	Percentage of subjects approved by the student
media_geral	Overall grade of the student in the course
media_Faltas	Average student absences in the course
cotista	Indicates whether the student entered the course by quota system

5) Dependent Variables: Accuracy (1), Recall (2), Precision (3) and F-Measure (4).

6) Experiment Design: After the preprocessing, which consisted in the removal of records that were very different from the average, 6672 instances were selected, which represent all undergraduate students of the institution, within the period

TABLE III
USED PARAMETERS BY ALGORITHM.

Algorithm	KNN	Random Forest	Naive Bayes	SVM	Decision Tree	MLP
n_neighbors	25	-	-	-	-	-
random_state	-	0	-	-	0	-
criterion	-	entropy -	-	-	gini	-
n_estimators	-	75 -	-	-	-	-
max_depth	-	10 -	-	-	None	-
n_jobs	-	-1 -	-	-	-	-
max_features	-	0.3 -	-	-	None	-
bootstrap	-	-	-	-	-	-
C	-	-	-	0.001	-	-
cache_size	-	-	-	200	-	-
class_weight	-	-	-	None	None	-
max_iter	-	-	-	-1	-	200
probability	-	-	-	False	-	-
random_state	-	-	-	None	-	1
shrinking	-	-	-	True	-	-
tol	-	-	-	0.001	-	0.0001
verbose	-	-	-	False	-	False
coef0	-	-	-	0.0	-	-
decision_function_shape	-	-	-	ovr	-	-
degree	-	-	-	3	-	-
gamma	-	-	-	1	-	-
kernel	-	-	-	poly	-	-
max_leaf_nodes	-	-	-	-	None	-
min_impurity_decrease	-	-	-	-	0.0	-
min_impurity_split	-	-	-	-	None	-
min_samples_leaf	-	-	-	-	1	-
min_samples_split	-	-	-	-	39	-
min_weight_fraction_leaf	-	-	-	-	0.0	-
presort	-	-	-	-	False	-
splitter	-	-	-	-	best	-
activation	-	-	-	-	-	logistic
alpha	-	-	1	-	-	1e-05
batch_size	-	-	-	-	-	auto
beta_1	-	-	-	-	-	0.9
beta_2	-	-	-	-	-	0.999
early_stopping	-	-	-	-	-	False
epsilon	-	-	-	-	-	1e-08
hidden_layer_sizes	-	-	-	-	-	(3, 2)
learning_rate	-	-	-	-	-	constant
learning_rate_init	-	-	-	-	-	0.002
momentum	-	-	-	-	-	0.9
shuffle	-	-	-	-	-	True
solver	-	-	-	-	-	adam

previously mentioned. Of the selected data, 3212 (48.1%) represents dropped out students and 3460 (51.9%) represent active students.

One of the metrics used in this work was the accuracy, which requires the balancing of class data. Since our base is already balanced [19], it was not necessary to plan the adoption of a balancing method.

The 10-fold Cross-validation approach was used, where the data are divided into 10 parts, maintaining their proportions. Thus, 10 tests are performed, in which part of the data is separated to be tested later and the others are used to be trained.

7) Instrumentation: For the data mining process, the Python language and its libraries was used, which has several machine learning algorithms that can be used to extract relevant information from a database. According to the 16th analysis software usage and data mining annuary [26], Python was considered the most used programming language by Data Mining and Big Data professional community.

Python has many reasons for attracting interest as a language for data analysis: it is open-source and free of cost, it has a varied set of libraries to work with several areas allowing performance comparison between the algorithms and presenting several resources for data analysis. In addition it offers a simple and objective syntax that allows the programmer to focus on the problem to be solved without worrying so much about details of implementations.

The data used for the analysis comes from SIGAA, which has the PostgreSQL as SGBD. An ETL (Extract, Transform and Load) was created to extrac, clean and load the data in a specific Data Warehouse, which is the basis for generation of knowledge models, taking into account the variables detailed in subsection V-A4.

VI. Experiment Operation

A. Preparation

It consisted of implementing the ETL implemented to load the Data Warehouse. The data were submitted to pre-processing, in which records with different values of the average (outlies) were removed. In this step, it was also done the transformation of some attributes, in which the "One Hot" approach was applied, consisting of representing a categorical variable of binary form. This process is represented in steps 2 and 3 of Figure 1.

B. Execution

It consisted in performing the classifying process in the data of the students, planned in section V-A6, for each selected mining algorithm, by using the dictionary discussed in subsection V-A4. Step 4 of Figure 1.

C. Data Validation

Four types of statistical tests, Shapiro-Wilk Test, Levene Test, Anova Test and Tukey Test were used as an aid to analysis, interpretation and validation - step 5 of Figure 1.

The Anova Test was used because it was necessary to compare more than two groups of values. Since this test has the assumptions that the distribution must be Normal and that there is homoscedasticity between the treatments (homogeneous variances) [9], the Shapiro-Wilk [30] Test was used for the Normality test and Levene's test [18] for the homoscedasticity test.

The Anova Test shows that at least one algorithm differs from the others, but it is not possible to say which one is more dissimilar. For this, the Tukey test was used, which according to Angels [2], allows to test any contrast, always, between two averages of treatments, being possible to verify which are statistically the same or different.

All statistical tests were done using the SPSS Tool - IBM [15].

VII. Results

After the execution of the algorithms using the 10-cross-validation approach, the results of the classifications were obtained. In Table IV and in Figure 2, the averages obtained by each algorithm with all the attributes are presented.

TABLE IV
Comparative of the metrics of the algorithms with all attributes.

Algorithms	Accuracy	Precision	Recall	F Measure
Knn	77,52%	77,56	77,51%	77,46%
Random Forest	80,63%	80,28%	80,33%	80,3%
Naive Bayes	77,13%	76,84%	77,45%	76,9%
SVM	81,01%	77,88%	78,02%	77,91%
Decision tree	77,62%	77,44%	77,7%	77,49%
MLP	79,15%	79,3%	79,25%	79,25%

By using the Random Forest algorithm for attribute selection, it is possible to notice that some have more relevance and others could be eliminated without influencing the results. The

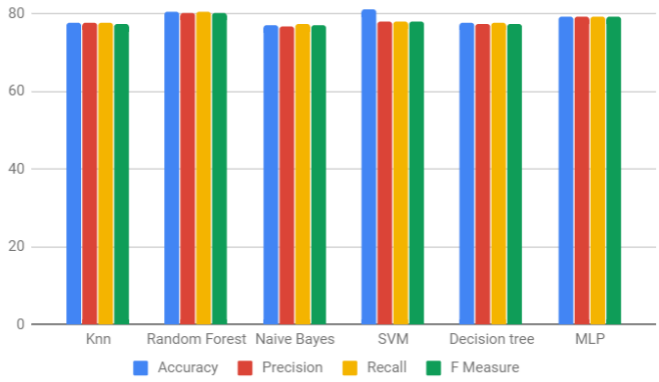


Fig. 2. Comparative chart of the metrics of the algorithms with all attributes.

4 attributes of greatest relevance had the following order, starting from the most relevant: "idade", "sexo", "qtd_ap1p" and "media_geral". The attributes that presented lower relevance were "qtd_rep_med_p" (average number of failed subjects per period), "qtd_per_cur" (percentage of finished courses) and "qtd_ap_med_p" (average number of courses approved perperiod).

This shows, for example, that the students' age and gender has influence on the problem and could be related with the responsabilities each of them have besides the studies. It also brings an alert for further analysis of the social areas of the institution, considering the gender and age group in the basic education that are being most affected. In addition, it is evident that the student's performance in the first period is quite relevant.

After selecting the attributes less relevant, the algorithms were executed again. The averages of each algorithm are presented in Table V and in Figure 3 below.

TABLE V
Comparative of the metrics of the algorithms with the selected attributes.

Algorithms	Accuracy	Precision	Recall	F-Measure
Knn	78,01	79,09	70,83	74,69
Random Forest	80,36	80,79	76,50	78,86
Naive Bayes	76,8	79,00	67,59	72,81
SVM	81,18	80,25	77,51	78,81
Decision tree	78,06	76,51	75,15	75,79
MLP	79,43	77,34	76,92	77,06

These results were used to respond the research question. The algorithms obtained distinct average accuracies and the SVM algorithm obtained the highest ones, followed by Random Forest, which achieved very simmilar averages and close to SVM's. However, it is not possible to make such assumptions without conclusive statistic evidence. For that reason, the Anova Test was applied to validate the hypotheses. At first, for having the assumptions of normality and homoscedasticity, the Shapiro-Wilk Test was carried out, followed by the Levene Test.

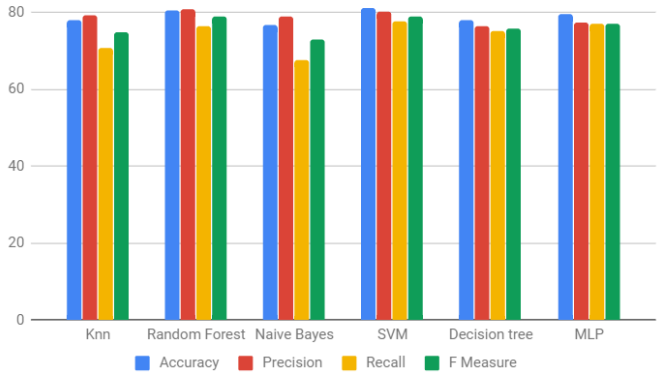


Fig. 3. Comparative chart of the metrics of the algorithms with the selected attributes.

A 0.05 significance level to the experiment was defined . When applying the Shapiro-Wilk Test for the normality analysis of the distribution of the data the p-values from Table VI were obtained. In which values above the level of significance were observed, leading to the conclusion that the distributions are normal.

TABLE VI
THE SHAPIRO-WILK TEST RESULTS FOR DATA NORMALITY ANALYSIS.

Algorithms	Accuracy	Precision	Recall	F-Measure
Knn	0,167	0,736	0,600	0,338
Random Forest	0,706	0,633	0,018	0,837
Naive Bayes	0,089	0,807	0,566	0,051
SVM	0,637	0,560	0,306	0,579
Decision tree	0,735	0,980	0,482	0,698
MLP	0,592	0,870	0,453	0,289

Then, the Levene Test was performed, obtaining the results presented in Table VII, that is, higher than the significance level adopted, validating the variances homogeneity assumption.

Once the assumptions were met, it was possible to apply the Anova test, in which p-values were significantly lower than the level of significance adopted, as can be seen in the table VII. In this way, the evidence of difference between the averages was confirmed, that is, the hypothesis (H_0) that the algorithms have the same accuracy was rejected within the context of the experiment.

TABLE VII
P-VALUES OF THE LEVENE AND ANOVA TESTS.

Metrics	Levene	Anova
Accuracy	0,807	$4,10^{-6}$
Precision	0,795	$3,02^{-6}$
Recall	0,978	$4,29^{-13}$
F-Measure	0,469	$4,01^{-9}$

The Anova Test evidence that at least one algorithm differs from the others, but it is not possible to affirm wich one. For this, the Tukey Test was used, because it allows to test any contrast between two averages treatment, making it possible to verify which are statistically different or equal, accordingly to [2].

Figure 4, following, presents the average accuracies grouped algorithms, forming four homogeneous groups. By analyzing the figure we, can see that th highest average belongs to the SVM algorithm (81.18%). However, considering the significance level of 5% it is possible to say that SVM and Random Forest obtained the same average accuracy. Naive Bayes achieved the lowest average (76.50%).

Algorithm	Subset for alpha = 0.05			
	1	2	3	4
Naive Bayes	76,80			
KNN		78,01		
Decision Tree		78,06	78,06	
MLP			79,43	
Random Forest				80,36
SVM				81,18
Sig.	1	0,865	0,101	0,597

Fig. 4. Values obtained by the Tukey's Test on Accuracy.

Similarly, the Figure 5 shows the result of the Tukey test of the F-Measure. To avoid repetition, the Tukey's test of precision and recall will be omitted in this work, since the F-Measure makes the combination of these metrics. The results show that the Random Forest, SVM and MLP algorithms presented similar averages for the F-Measure, being 78.86%, 78.81% and 75.99%, respectively. Naive Bayes had the lowest average with 72.81%.

Algorithm	Subset for alpha = 0.05			
	1	2	3	4
Naive Bayes	72,81			
KNN	74,69	74,69		
Decision Tree	75,79	75,79	75,79	
MLP		75,99	75,99	75,99
SVM			78,81	78,81
Random Forest				78,86
Sig.	0,056	0,802	0,051	0,073

Fig. 5. Values obtained by the Tukey Test's on F-Measure.

A. *Threats to Validity*

Although the results of the experiment were satisfactory, it presents threats to its validity that should be commented on.

Threats to internal validity: The current academic system has been present in the Institution since 2017, which inherited the basis of the academic system legacy, with several inconsistent information, mainly until the middle of 2007. This threat was mitigated with the accomplishment of the cleaning process that decreased the likelihood of using incorrect older information.

Threats to construction validity: In the experiment of this article, the institution did not possess any very relevant information regarding dropout, like, for example, socioeconomic data of the students and their relatives. The inclusion of this information can influence the performance of the algorithms, increasing their efficiency. To mitigate this threat and increase the Decision Support System yet to be developed, we will suggest that this information be gathered by the institution and taken into account in future work.

VIII. Final Considerations

The main contribution of this work was to evaluate six major algorithms most commonly used in the context of EDM in terms of Accuracy, Precision, Recall and F-Measure for identify identifying the factors that influence school dropout.

The work was consolidated with the conduction of a controlled experiment in which the results showed that there are significant differences between the algorithms used, and that the algorithms Random Forest and SVM have stood out in this context, possessing, statistically, similar Accuracy (80.36%, 81.18%), Precision (80.79%, 80.25%), Recall (76.50%, 77.51%) and F-measure (78.86%, 78.81%) averages. The results shows evidence of significant differences between the algorithms, and that although the Random Forest and the SVM has the best metrics evaluated, its results are statistically similar with MLP results.

Besides that, the published works on the subject have some scientific gaps if we consider that there was no rigorous validation of the results, allowing a more assertive combination of experimental evidence. In this context, this paper validates its results and confirms some previous evidence found in the works described in [22] e [10].

Finally, as future work, we intend to analyze the algorithms in other levels of education, as well as add other types of variables for analysis, such as socioeconomic information. In addition, we intend to develop a predictive system for teaching management, which can help in decision making process to combat school dropout and school retention.

References

[1] A. M. Ahmed, A. Rizaner, and A. H. Ulusoy. Using data mining to predict instructor performance. *Procedia Computer Science*, 102:137–142, 2016.

[2] A. Anjos. Análise de variância. *Universidade Federal do Paraná, Departamento de Estatística - UFPR, Curitiba*, page Capítulo 7, 2009.

[3] V. R. Basili and D. M. Weiss. A methodology for collecting valid software engineering data. Technical report, NAVAL RESEARCH LAB WASHINGTON DC, 1983.

[4] L. Breiman. Machine learning. *Kluwer Academic Publishers*, pages 5–32, 2001.

[5] L. Breiman, J. Friedman, R. Olshen, and C. Stone. *Classication and regression trees*. Monterey, CA: Wadsworth and Brooks, 1984.

[6] G. O. Campos, A. Zimek, J. Sander, R. J. G. B. Campello, B. Micenková, E. Schubert, I. Assent, and M. E. Houle. On the evaluation of unsupervised outlier detection: measures, datasets, and an empirical study. *Data Mining and Knowledge Discovery*, 30(4):891–927, Jul 2016.

[7] C. Cortes and V. Vapnik. Support-vector networks. machine learning, 20. pages 273–297, 1995.

[8] I. N. de Estudos e Pesquisas Educacionais Anísio Teixeira – Inep. Censo da educação superior 2015. 2015.

[9] A. Field. *Descobrindo a estatística usando o SPSS*. 2.ed. Porto Alegre: Artned, 2009.

[10] D. G., P. M., and V. J. Predicting students drop out: A case study. *Proceedings of the International Conference on Educational Data Mining*, pages 41–50, 2009.

[11] F. Gorunescu. *Data Mining: Concepts, models and techniques*, volume 12. Springer Science & Business Media, 2011.

[12] J. Han, J. Pei, and M. Kamber. *Data mining: concepts and techniques*. Elsevier, 2011.

[13] D. J. Hand, H. Mannila, and P. Smyth. *Principles of data mining*. MIT press, 2001.

[14] N. Iam-On and T. Boongoen. Improved student dropout prediction in thai university using ensemble of mixed-type data clusterings. *International Journal of Machine Learning and Cybernetics*, 8(2):497–510, 2017.

[15] IBM. Spss. *IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp*, 2017.

[16] N. Juristo and A. Moreno. Software engineering experimentation. 2001.

[17] G. Kantorski, E. G. Flores, J. Schmitt, I. Hoffmann, and F. Barbosa. Predição da evasão em cursos de graduação em instituições públicas. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 27, page 906, 2016.

[18] H. Levene. Robust tests for equality of variances. *International Journal of Machine Learning and Cybernetics*, pages 278–292, 1960.

[19] E. Machado and L. Marcelo. Um estudo de limpeza em base de dados desbalanceada e com sobreposição de classes. *XXVII Congresso da Sociedade Brasileira de Computação*, 2007.

[20] R. D. Machado, E. O. B. Nara, J. N. C. Schreiber, and G. A. Schwingel. Estudo bibliométrico em mineração de dados e evasão escolar. *XI Congresso Nacional de Excelência em Gestão*, 2015.

[21] L. M. B. Manhães, S. Cruz, R. J. M. Costa, J. Zavaleta, and G. Zimbrão. Identificação dos fatores que influenciam a evasão em cursos de graduação através de sistemas baseados em mineração de dados: Uma abordagem quantitativa. *Anais do VIII Simpósio Brasileiro de Sistemas de Informação, São Paulo*, 2012.

[22] C. Márquez-Vera, C. R. Morales, and S. V. Soto. Predicting school failure and dropout by using data mining techniques. *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje*, 8(1):7–14, 2013.

[23] E. Mundstock. Introdução à análise estatística utilizando o spss 13.0. cadernos de matemática e estatística série b. 2006.

[24] A. Nürnberger, W. Pedrycz, and R. Kruse. *Handbook of data mining and knowledge discovery. Chapter data mining tasks and Methods: Classication: Neural network approaches*. New York, NY, USA: Oxford University Press, 2002.

[25] J. G. d. Oliveira Júnior et al. Identificação de padrões para a análise da evasão em cursos de graduação usando mineração de dados educacionais. Master's thesis, Universidade Tecnológica Federal do Paraná, 2015.

[26] K. S. Poll. Analytics, data mining software used. *https://www.kdnuggets.com/polls/2015/analytics-data-mining-data-science-software-used.html*, 2015.

[27] A. Pradeep, S. Das, and J. J. Kizhekkethottam. Students dropout factor prediction using edm techniques. In *Soft-Computing and Networks Security (ICSNS), 2015 International Conference on*, pages 1–7. IEEE, 2015.

[28] S. J. Russell and P. Norvig. *Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited,, 2016.

[29] A. J. Severino. *Metodologia do trabalho científico*. Cortez editora, 2017.

[30] S. Shapiro and M. Wilk. An analysis of variance test for normality (complete samples). *International Journal of Machine Learning and Cybernetics*, 52:591–611, 1965.

[31] D. F. Silva and G. E. de Almeida Prado Alves Batista. Uma comparação experimental de métodos de imputação de valores desconhecidos. *ICMC - Instituto de Ciências Matemáticas e de Computação, São Paulo*, 2009.

[32] R. L. L. Silva Filho, P. R. Motejunas, O. Hipólito, and M. Lobo. A evasão no ensino superior brasileiro. *Cadernos de pesquisa*, 37(132):641–659, 2007.

[33] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén. *Experimentation in software engineering*. Springer Science & Business Media, 2012.

[34] X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, S. Y. Philip, et al. Top 10 algorithms in data mining. *Knowledge and information systems*, 14(1):1–37, 2008.

12th International Workshop on Computational Optimization

MANY 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
- **Stoean, Catalin,** University of Craiova, Romania
- **Wang, Yifei**
- **Zilinskas, Antanas,** Vilnius University, Lithuania

Communication Papers of the Federated Conference on Computer Science and Information Systems pp. 13–19
DOI: 10.15439/2019F95
ISSN 2300-5963 ACSIS, Vol. 20

Selecting representatives

László Aszalós
Faculty of Informatics at University of Debrecen,
26 Kassai út H4028 Debrecen , Hungary
Email: aszalos.laszlo@inf.unideb.hu

Dávid Nagy
Faculty of Informatics at University of Debrecen,
26 Kassai út H4028 Debrecen , Hungary
Email: nagy.david@inf.unideb.hu

***Abstract*—We use representatives to reduce complexity in many areas of life. Clusters are often replaced with their centre, and then these representatives are used to classify new objects. If the objects are described as a vector of real numbers, then the centre can be easily calculated. However, this method is unusable if only a similarity relation is given instead of coordinates of the object or the distances between the objects. Google can filter and rank relevant pages for a particular question; and here we follow a similar approach. The difference is that we have an undirected graph while the PageRank algorithm uses a directed one. In this article we show what conditions we set for our own ranking system. Following the description of the details of this method we demonstrate that it satisfies our criteria and how it selects the (mathematically proven) most typical elements of each cluster. Finally, we apply this method on several partitions of the natural numbers and on non-transitive tolerance relations to present the representatives of the numbers.**

I. Introduction

INSTEAD of examining the entire population, polls usually only survey a small sample. This can be done because the results obtained are very close to what we would get by examining the entire population. However, the sample should be selected carefully. Almost everyone knows the concept of the *representative sample*, but only few know exactly what it means. Many people think that the larger the sample, the better; which is not true. The sample is representative in some respects, i.e., the specific properties are as similar in the sample as in the entire population. The sample can be representative in one aspect, while not representative in another. There are various standard methods for determining the sample.

If the population is significantly inhomogeneous, i.e. it has high variability according to the survey, then the stratified (random) sampling is used. In this case, the population is divided into several sub-populations (strata), where these sub-populations are homogeneous according to the examined criteria. From a homogeneous strata, we can randomly select the individuals to be sampled (i.e., the representative of the group), typically in proportion to the size of the group.

If we can represent an object with a vector of numbers, we can consider the difference of vectors belonging to each object, where this difference/distance usually meets the requirements of metrics. Using this distance function, many clustering methods have been developed over the last sixty years. The most well-known *k-means method* replaces a cluster with its centre (one representative). The *k-medoids algorithm* is a version of this k-means method, and it replaces the cluster with the sample element closest to the cluster centre. The *CURE method* (clustering using representatives) goes one step further, replacing non-ellipsoid clusters with maximum c sample elements. The most common use of k-means (or its enhancements) is the *k-nearest neighbours* (k-NN) classification algorithm, where newly added objects must be categorized into an existing cluster/class. Since comparing the new elements with all stored elements in a large database is a time costly task, by replacing the elements of the clusters with some of their representatives we can significantly reduce the complexity of the classification of new elements.

Polls can not ask too many questions from a person because their patience is finite. However, there are cases where we leave behind a lot of information, think for example our medical cards, our data stored at different kinds of service providers, or our digital footprint on the social network. In these cases, it is not worth transforming this information into a unified form in order to be able to define the differences between the data of objects. It is much easier to directly decide for two given objects whether they are similar or not.

In this article, we present a mathematical method which—having an existing partition and similarity relation–determines which is the most typical object in a given cluster, i.e. which one can be considered *representative*. We assign a real number, a *rank* to each of these objects, and the highest-ranking object in each cluster becomes the representative of the cluster.

In the next section, we present the requirements we expect from the rank of the objects. In the Section III we present the power method and how it can be used for our purposes. Next we demonstrate the results of our method through two special relations, and how matches our expectations. Finally we conclude our results.

II. The properties of ranking

In the following, we identify how to describe relations using signed graphs. Each element in a relation will be a vertex of the graph and two vertices are connected with an edge if and only if their two corresponding elements are in relation.

In case of graphs, we can speak of the distance between two vertices (as the shortest path between the two vertices), but it carries much less information than the difference of two large vectors. Therefore, the similarity information should

The publication is supported by the EFOP-3.6.1-16-2016-00022 and EFOP-3.6.3-VEKOP-16-2017-00002 projects. These projects are co-financed by the European Union and the European Social Fund.

©2019, PTI

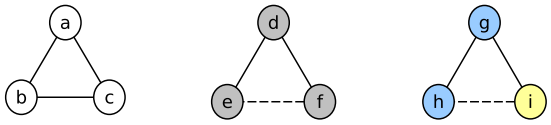


Fig. 1. Simple ranking problems

already be included in the graph, so the graph will correspond to a similarity relation. As we usually have partial similarity relations in practice, we will have edges in the graph that denote the similarity and we will have edges that denote the dissimilarity. The partiality is represented by missing edges.

For example, links between individual websites or citations between scientific articles define a directed graph, i.e. a partial similarity relation, but there is no representation of *dislike* i.e. dissimilarity.

Google's PageRank algorithm [2] is a great example of a ranking system on directed graphs. Considering the web pages (vertices) and the links between them (edges) as a directed graph, the boundary distribution of the random walk on the graph gives the rank of each page. For example, if the web page p is more likely to be accessed than the web page q, then the rank of p will be higher than that of q and will therefore be ranked higher in the hit list. Here, if a page with a low rank refers to a page with a high rank, or a novice author refers to a well-known author in his article, it raises the rank of the page/author to a higher rank, but—through a non-symmetrical relation—this reference has no effect on the rank of the page/author with the lower rank.

However, if the graph is not directed, the edge between the two vertices will affect the rank of both vertices. Since the similarity relation is a tolerance relation (reflexive, symmetric, but not necessarily transitive), the associated graph is not directed.

Let's see our (naive) expectations of a ranking method. In Fig. 1 on the left, we have three vertices (a, b and c) in a common cluster. In this figure, a cluster is represented by vertices of the same colour, while the similarity of vertices is denoted by a solid, and the difference by a dashed line. Because in this sub-graph each element is similar to each other, we expect the same rankings due to symmetry. In the middle graph of Fig. 1—where d, e and f are in a common cluster—a difference appears. This graph is called a *minimal frustrated graph*, because there is no such partition of vertices where similar elements are common, and different elements are clustered separately. In this graph, vertex d has only similar vertices, while vertices e and f both have similar and also different vertices. The fact that an object differs from an object in its own cluster reduces the rank of the object/vertex and thus the chance of being a representative of the cluster. Conversely, if an object similar to an object in its own cluster, then this increases its rank. Based on these, this cluster will be represented by vertex d because it will have the highest rank. Moreover—according to the symmetry—the rank of vertices e and f should be the same.

Finally, take the graph on the right side of the Fig. 1. Here, the vertices were divided into two clusters: $\{g,\ h\}$ and $\{i\}$. The fact that the vertices of g and i are similar, but are found in different clusters also reduces the rank of both vertices, because similarity to vertices in others clusters means *deviation* from the idealized characteristics of the group. The vertex h is similar to vertex g which belongs to h's cluster, and h is different from the vertex i belongs to another cluster. This latter also raises the rank of the vertex h and hence h becomes the representative of its own cluster. In the other cluster, the only vertex will be the representative.

Based on the examples above, the similar objects of the same cluster and dissimilar objects of other clusters can be called the *fosterer* of the object, while the similar objects of different clusters, and dissimilar objects of the same cluster can be called the *adversary* of the object. The fosterer objects help an object to become a representative, while the adversary objects prevent it this from happening.

Let's summarise what we would expect from the rankings.

- Be symmetric, that is, if two vertices have the same number of vertices of the same rank in the same type relation (fosterer or adversary), then their rank is the same.
- The rank of a particular item is immediately raised if:
  - one of its fosterer object increases in rank, or
  - one of its adversary object falls in rank, or
  - a new fosterer object appears.
- The rank of a particular item is immediately reduced if:
  - one of its adversary object increases in rank, or
  - one of its fosterer object falls in rank, or
  - a new adversary object appears.
- It does not directly change the rank of a particular item if:
  - another object that is not compared to it or is incomparable appears in any cluster, or
  - the rank of such an object changes.

We were mindful of our usage of the word *immediately*. If a new object that is dissimilar to the current object, but similar to another object of the cluster, is added to this cluster, then it raises the rank of objects that are similar to it. This will have a ripple effect on objects that are similar to objects are similar to the new object, and so on. Therefore, if we think of representing this with an algorithm, we need an iterative method that will escalate these effects step by step. On the other hand, since almost every objects are related to every objects, we need to treat the rank of all objects altogether.

III. Our ranking method

Let V denote the set of objects/vertices, and for simplicity, denote the objects with numbers: $V = \{1, 2, \ldots, n\}$. The set of the clusters means a partition. This partition is interpreted as a function—denoted by p—that assigns a number to each object, so $p : V \to \mathbb{N}$. The objects i and j are in the same cluster, if $p(i) = p(j)$; and they are in different clusters, if

$p(i) \neq p(j)$. Our similarity relation is a (possibly partial) tolerance relation T—that is reflexive, symmetric, but not necessarily transitive. In Section IV we present the numerical results of our ranking method through two partial tolerance relations.

A. Social ranking

We use similar approach as PageRank or various evaluation sites (accommodations, restaurants, marketplaces), where the rankings of individual websites, hotels, restaurants are summed up by aggregating individual ratings.

We could take the rank of each object as the difference between the number of fosterer and adversary objects, but we want to introduce a more sophisticated method. We think that there is a significant difference between the the cases where the adversary object is the representative of the other cluster, or it just a marginal object there. In the former case it decreases the rank of the current object to a greater extent. We find that a value proportional to the rank of the adversary or fosterer objects defines a a good amount by which we can decrease or increase the rank of the current object.

The rank of object i is determined by its relation to all objects. Let a_{ij} indicate the relation (fosterer or adversary) between objects i and j. We want to consider each object with its rank as weights, and so we get the following relation: $r_i = \sum_j a_{ij} \cdot r_j$ for all $i \in V$, which we want to solve for values r_i. These equations combine to $R = AR$ in matrix notation.

So that we do not have to use different values a_{ij} from task to task—depending on its size—let's introduce a constant c to normalise the values a_{ij}. This changes the previous relation as follows $r_i = \sum_j (c \cdot a_{ij}) \cdot r_j$, but it can be transformed to $r_i = c \cdot \sum_j a_{ij} \cdot r_j$, which gives us $R = cAR$. If we take the reciprocal of c to be denoted by λ, we get a known relation: $AR = \lambda R$, i.e. R is an eigenvector of the matrix A.

Taking into account the ideas from the previous chapter, the normalized value a_{ij} should be 1 for fosterer objects, -1 for adversary objects, and 0 for all other cases (where the tolerance relation is partial).

According to the much cited example, a bald man does not resemble a hairy man, although an uprooted hairline does not change a person. As a person can have up to two million hairlines, the linear model of this example is reduced to only four states. Here 1 (bald) and 4 (hairy) correspond the two end states, while 2 and 3 are two intermediate states. 1 and 2 are in a common cluster, and so are 3 and 4. Fig. 2 shows two graphs demonstrating this problem, where we denote clusters by assigning different colours to the vertices. The similarities (denoted by solid lines) are the same in both cases, but the dissimilarities (denoted by dashed lines) have changed: dissimilarity appears between the second neighbours in the latter case. Hence the first graph/relation is partial, and the second is total.

Two matrices based on the relations and partitions for each graph are:

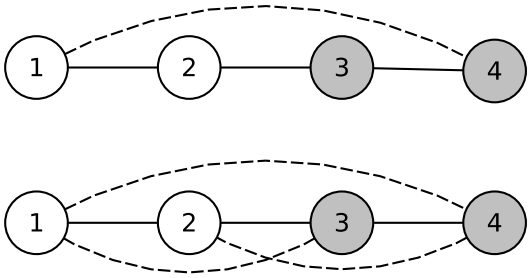


Fig. 2. Two simple symmetric linear model

$$A_1 = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & -1 & 0 \\ 0 & -1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix} \text{ and } A_2 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

The only adversary relation is between 2 and 3, because they are similar, but are in different clusters. In the first case 1-3 and 2-4 are not comparable, so the corresponding values of the matrix are 0. In the second case these pairs of objects are dissimilar, but they are in different clusters, so these are fosterer relations.

Calculating the eigenvalues and eigenvectors for the first graph, we get:

r_1	r_2	r_3	r_4	λ
0.7071	0.7071	0.3344	0.1368	2.4142
0.5000	-0.5000	0.6770	-0.5873	-0.4142
0.0000	0.0000	-0.6230	-0.6938	2.4132
0.5000	-0.5000	-0.2041	0.3939	-0.4142

Our graph is symmetric, so we would expect, that $r_1 = r_4$ and $r_2 = r_3$. Unfortunately, none of the eigenvectors satisfy this. Therefore, we need to look for another method!

B. Power-method

The algorithm of von Mises [5] for a diagonal matrix A results in the biggest eigenvalue (with the highest absolute value), and the corresponding eigenvector. The method starts with an arbitrary vector R_0 that in our case should be $\mathbf{1} = (1, \ldots, 1)^T$. Then R_{k+1} is determined as follows: the rank vector R_k is multiplied by the matrix A and normalized as shown by (1).

$$R_{k+1} \leftarrow \frac{AR_k}{||AR_k||} \tag{1}$$

Unfortunately, this iteration is converges slowly, but it is easy to use even for large sparse matrices; which is why it is used in PageRank implementation. If $R_i \approx R_{i+1}$, the method is stopped and the values in the vector R_i are considered the rank of the objects. If matrix A has an eigenvalue that is strictly greater in magnitude than its other eigenvalues, then R_i converges.

Algorithm 1 Python implementation of our ranking method

```
def power_method(A, eps = 1e-9):
    B = A
    R = np.ones((len(A),))
    B2 = B@B
    B2 /= m = np.max(B2@R)
    while np.linalg.norm(B@R - B2@R) > eps:
        B, B2 = B2, B2@B2
        B2 /= np.max(B2@R)
    return B2
```

If this method is applied to the matrices shown in Fig. 2, then we get $R = (1, 0.414, 0.414, 1)^T$ and $R = (1, 0.618, 0.618, 1)^T$, where $r_1 = r_4$ and $r_2 = r_3$ as we expected. These values also fit to our naive ideas: for the first graph, for object 1 both objects 2 and 4 are fosterers, while for object 2 object 1 is a fosterer and object 2 is an adversary. So the expected relation $r_1 > r_2$ is fulfilled. In the second case, when there are more fosterer relations, the rank gained is also higher for objects 2 and 3.

Summarising the gives:

$$R_{k+1} \leftarrow \frac{AR_k}{||AR_k||} \approx \frac{A^{k+1}\mathbf{1}}{||A^{k+1}\mathbf{1}||}$$

If k is a power of 2, then we can then calculate the same values by repeated squaring using the following recurrence relation:

$$B_1 \leftarrow A \text{ and } B_{i+1} \leftarrow \frac{B_i B_i}{||B_i B_i||} \tag{2}$$

If $B_i\mathbf{1} \approx B_{i+1}\mathbf{1}$, then let $R \leftarrow B_i\mathbf{1}$. Not surprisingly, the two calculations give the same result.

Based on these, it is not difficult to write the ranking program in Python using the services of the Numpy package (Algorithm 1). Remark, that for Numpy, the operator `@` is the matrix multiplication operation.

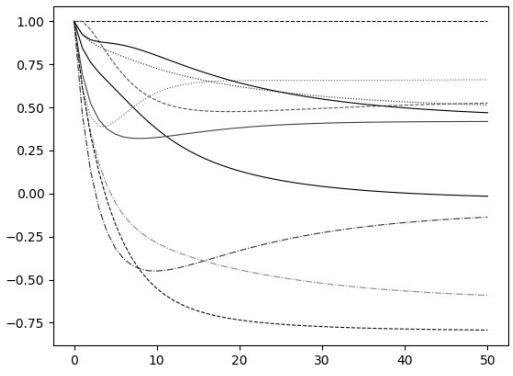


Fig. 3. Changes in the rank of the objects in a random tolerance relation

Fig. 3 shows how each value r_i changes for a random matrix A. Here, the axis x represents the number of applications of (2), while the axis y represents the current values of the ranks. Due to normalisation, the highest rank is always 1, but as we can see from the chart, the rank of the objects changes from time to time. The algorithm is terminated when the ranks cease changing.

IV. Ranks of numbers

As there is no standard similarity relations for larger databases, the various clustering/classification methods are usually tested on random graphs [6]. On the other hand, we spent a long time looking for a tolerance relation, that is **not** an equivalence relation, freely scalable, easy to understand for the reader and enables reproducible computations. Finally, we found the *common divisor* as the basis of similarity: let two numbers be similar if they have a non-trivial common divisor, i.e. $\gcd(a, b) > 1$, where $a,\ b \in \mathbb{N}^+$. Then $4T6$ and $6T9$ are fulfilled (the common divisors are 2 and 3), but $4T9$ is not, so this relation is not transitive. If we are interested in the correlation clustering of numbers $1, \ldots, n$, it can be easily formulated if $n < 111\,546\,435$, otherwise the situation become complicated [3].

In the following, the rank of each element is determined by using the optimal clustering of the set of numbers, except in the first case, where we place the numbers $1, \ldots, 12$ in a common cluster.

As we have a single cluster in the equation (2), we can replace the matrix A in the calculation with the matrix of the relation T in Fig. 4. In this matrix T, the relation between the numbers i and j is given by the j^{th} number from the i^{th} row: similar numbers are denoted by 1, dissimilar numbers by -1. 12 and 6 are similar to the multiples of 2 and 3, these are eight numbers including themselves, and different from four of them (1, 5, 7, and 11). The powers of 2 are similar to every even number (six numbers) and different from all odd number (six numbers). The powers of 3 (itself and 9) are similar to four numbers and different from eight numbers. 5 is similar to its duplicate, but there is no such number for 7 and 11 and for 1. Because in each case the numbers mentioned together are similar to the same numbers, so—according to the symmetry—their rank is the same (Table I/A). In this table the ranks are rounded to the closest hundreds, to fit on the page.

If we do the same calculation for numbers $1, \ldots, 100$ (Table I/B) then the numbers will typically increase. There are many more even numbers, which will increase the rank of even numbers, but they have the opposite effect on the rank of powers of 3. In this set, there are numbers similar to 7 or 11, so their rank increases; and there are more such numbers for 7, so its rank grows more.

A. Using optimal partition

Consider the partition of natural numbers obtained by correlation clustering [3]. The largest such cluster is the set of even numbers. This is followed by a set of odd numbers divisible by 3; next the set of numbers which are divisible

TABLE I
RANKING NUMBERS...

A) $1, \ldots, 12$ in a common cluster.

1	2	3	4	5	6	7	8	9	10	11	12
-0.62	0.97	0.06	0.97	-0.32	1.00	-0.62	0.97	0.06	0.89	-0.62	1.00

B) $1, \ldots, 100$ in a common cluster.

1	2	3	4	5	6	7	8	9	10	11	12	...
-0.55	0.99	-0.03	0.99	-0.28	1.00	-0.37	0.99	-0.03	0.95	-0.44	1.00	...

C) $1, \ldots, 12$ using optimal partition.

C_1	C_2						C_3		C_5	C_7	C_{11}
1	2	4	6	8	10	12	3	9	5	7	11
1.00	1.00	1.00	0.73	1.00	0.84	0.73	0.73	0.73	0.84	1.00	1.00

D) $1, \ldots, 15$ using optimal partition.

C_1	C_2							C_3			C_5	C_7	C_{11}	C_{13}
1	2	4	6	8	10	12	14	3	9	15	5	7	11	13
1.00	1.00	1.00	0.67	1	0.79	0.67	0.86	0.79	0.79	0.54	0.79	0.86	1.00	1.00

E) $1, \ldots, 12$ using common cluster and the weakened relation.

1	2	3	4	5	6	7	8	9	10	11	12
-0.53	1.00	0.06	0.83	-0.34	0.68	-0.53	0.70	-0.17	0.42	-0.53	0.81

F) $1, \ldots, 12$ using optimal partition and the weakened relation.

C_1	C_2						C_3		C_5	C_7	C_{11}
1	2	4	6	8	10	12	3	9	5	7	11
1.00	1.00	0.89	0.53	0.83	0.53	0.62	0.77	0.88	0.89	1.00	1.00

$$T = \begin{pmatrix}
1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\
-1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\
-1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 \\
-1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\
-1 & -1 & -1 & -1 & 1 & -1 & -1 & -1 & -1 & 1 & -1 & -1 \\
-1 & 1 & 1 & 1 & -1 & 1 & -1 & 1 & 1 & 1 & -1 & 1 \\
-1 & -1 & -1 & -1 & -1 & -1 & 1 & -1 & -1 & -1 & -1 & -1 \\
-1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\
-1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\
-1 & 1 & -1 & 1 & 1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\
-1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & -1 & -1 \\
-1 & 1 & 1 & 1 & -1 & 1 & -1 & 1 & 1 & 1 & -1 & 1
\end{pmatrix},$$

$$A = \begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & -1 & 1 & 1 & 1 & 1 & 1 & -1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & 1 & 1 \\
1 & 1 & -1 & 1 & 1 & 1 & 1 & 1 & -1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & -1 & 1 & 1 & 1 & 1 & 1 & -1 \\
1 & 1 & 1 & 1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & -1 & 1 & 1 & 1 & 1 & 1 & -1 & 1 & 1 & 1
\end{pmatrix}$$

Fig. 4. Similarity relation on numbers $1, \ldots, 12$, and the suitable matrix based on the optimal partition.

by 5, but not by 3 or 2, and so on. This can be formulated as $C_2 = \{x \in U : 2|x\}$, $C_3 = \{x \in U : 3|x\} \backslash C_2$, $C_5 = \{x \in U : 5|x\} \backslash C_2 \backslash C_3$, ... Of course, each prime number has one (possibly empty) cluster.

If we consider the fosterer and adversary objects based on the the optimal partition and the similarity relation T in Fig. 4, then we get the matrix A presented in Fig. 4. The partition mentioned above is optimal because it minimises the number of negative numbers in the matrix A. This, of course, also has an impact on rankings. While keeping the numbers in a common cluster, multiple ranks were negative due to dissimilarities, by using the optimal partition each cluster as a cluster by applying similarity (Table I/C). For singletons (containing big primes and 1) the rank 1.0 is natural, as there are no similar numbers. In the set of even numbers, the rank of the powers of two will also be 1.0, as they are similar to all even numbers, and the C_2 cluster has only even numbers, and all even numbers are here. Numbers with other divisors will have similar numbers in other clusters. The more prime divisors a number has, the more clusters contain similar numbers, so its rank will be reduced. In the cluster C_3 the rank of the powers of three will be the highest, but it will not reach level 1.0, because there are nearly as many numbers that are dividable by 3—that is, similar—in the cluster C_2, as in the cluster C_3.

If we apply our method on numbers $1, \ldots, 15$ instead of numbers $1, \ldots, 12$, then the ranks change (Table I/D). The rank of 7 fell, as its adversary number (14) appeared. As number 15 is adversary for both 5 and 10, it reduces the rank of both of them. The reader may be wondering how these ranks look when we have more numbers, e.g. $1 \ldots, 1000$ (Table II, column r). Perhaps it is clear from the above that in case of an optimal clustering, the representatives of the individual clusters come from the powers of primes. Each power of a prime is given the same rank because it is completely symmetrical in terms of similarity.

B. Weakening the tolerance relation

Let's see what happens if we expect one of the numbers to be the *divisor* of the other number instead of the existence of a real common divisor. (Remember, the relation must be symmetrical.) The partial relation here is made from a complete relation. We suggest to the reader to compare the first matrices in Fig. 4 and 5. If two numbers were dissimilar at the original tolerance relation, they will be dissimilar at the

TABLE II
RANG OF NUMBERS $1, \ldots, 1000$ USING THE OPTIMAL PARTITION

x	$r(x)$	$r'(x)$
1	1.0	1.0
2	1.0	1.0
3	0.71587172	0.79465596
5	0.78447065	0.81182099
7	0.82770503	0.84160798
11	0.88038405	0.88841938
13	0.89607481	0.90279018
17	0.91498232	0.91946078
4	1.0	0.8102978
6	0.75571551	0.55292589
8	1.0	0.71548803
10	0.85121315	0.5761841

$$T' = \begin{pmatrix}
1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\
-1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\
-1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 & -1 & -1 & 1 \\
-1 & 1 & -1 & 1 & -1 & 0 & -1 & 1 & -1 & 0 & -1 & 1 \\
-1 & -1 & -1 & -1 & 1 & -1 & -1 & -1 & -1 & 1 & -1 & -1 \\
-1 & 1 & 1 & 0 & -1 & 1 & -1 & 0 & 0 & 0 & -1 & 1 \\
-1 & -1 & -1 & -1 & -1 & -1 & 1 & -1 & -1 & -1 & -1 & -1 \\
-1 & 1 & -1 & 1 & -1 & 0 & -1 & 1 & -1 & 0 & -1 & 0 \\
-1 & -1 & 1 & -1 & -1 & 0 & -1 & -1 & 1 & -1 & -1 & 0 \\
-1 & 1 & -1 & 0 & 1 & 0 & -1 & 0 & -1 & 1 & -1 & 0 \\
-1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & -1 \\
-1 & 1 & 1 & 1 & -1 & 1 & -1 & 0 & 0 & 0 & -1 & 1
\end{pmatrix}$$

$$A = \begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & -1 & 1 & 1 & 1 & 1 & 1 & -1 \\
1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & 1 & 1 \\
1 & 1 & -1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\
1 & 1 & 1 & 0 & -1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & -1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1
\end{pmatrix}$$

Fig. 5. Weakened similarity relation on numbers $1, \ldots, 12$, and the suitable matrix based on the optimal partition.

weakened relation, too. Moreover for example, numbers 4 and 6 were similar before, but not anymore; so the relation holds less often.

At first glance in Table I/E the ranks of 2, 4, 8 are different, as well as ranks of 3 and 9. Once we have only one cluster, it is sufficient to count how many positive and negative values are in each row of the first matrix in Fig. 5. The number 2 is similar to every even number, that is, to every object in its own cluster C_2, and there is no other similar numbers anywhere else. The number 4 is similar to half of the numbers of its own cluster, but not dissimilar to any numbers in this cluster. The number 8 is similar to quarter of the numbers of its own cluster, etc. The fact that we replaced some ones with zeros, the symmetry disappeares. We remark that the numbers 1, 7 and 11 are dissimilar from all the other numbers in this case, so now they have the same (negative) rank.

Let's see what happens, if we take our previous optimal partition (Table I/F). The previous asymmetry remains. The primes have the highest rank, and the powers of primes have lower ranks. To see this tendency, let's see the outcome of ranking $1, \ldots, 1000$ (Table II, column of r').

If we apply the optimal partition for tolerance relation T to the weakened tolerance relation T' (Table I/F), the number of negative numbers in matrix A will also be significantly reduced. However, because of the change in the relation, the symmetry within the clusters is severely damaged, which also affects the rankings. Therefore, when some ranks weaken in the C_2 cluster, this may affect elements in other clusters. If there exists fosterer elements in other clusters, these will in turn increase in rank.

Which partition could be optimal for this weakened relation? Let us take the number of similar and dissimilar numbers in the clusters according to some given number. If the difference of these numbers not in their own cluster is maximal, by moving the actual number to the maximal cluster, we get a better partition. So now we wish to examine, whether we improve the optimal partition. Let $V = \{1, \ldots, 1000\}$, and take the number 75 as an example, which is $3 \cdot 5^2$. With the original relation T, C_2 contains 267 numbers are similar to it (which are divisible by 3 or 5), and 233 which are dissimilar. C_3 contains 167 similar numbers (all of them divisible by 3) and C_5 contains 67 similar numbers. At a weakened relation T', the previously dissimilar numbers remain dissimilar, and in C_2 there are no divisors of 75, just its even multiples. These by definition are similar to it, so C_2 contains 6 members similar to 75. In C_3 we have three divisors of 75 (including itself), as well as 6 of its odd multiples, so C_3 contains 9 similar numbers. In C_5 we have two divisors of 75 and it does not contain any of its multiples (because any multiple needs to have a prime divisor 3, so it would either be in C_3, or in C_2 if the number is even). If we take the difference of similar and dissimilar numbers for each clusters, this will be maximal in the number's own cluster, as here for C_3. This means that this kind of partition is stable for T', and with a very high probability it is the optimal partition, but this needs to be proved.

V. CONCLUSION AND FURTHER WORK

In this article, we presented a method that assigns a rank to each object by using a similarity relation and a partition of objects. In a special case, the method can be used for a single cluster, as shown by two examples. The similarity relation does not have to be complete, the method also works for partial relations. Based on our tests, using successive squaring to apply our method to a thousand objects, it is enough to use around dozens of squaring matrices. Since the tools used (matrix multiplication, norm calculation) are available in standard mathematical packages in almost all programming languages, this method is widely applicable.

Our future plans include a method to test real-life data, for example, in case of wines described by their chemical characteristics [1], we want to search for a typical wine—i.e. representative—of some wine region. Similarly we are planning to apply this method to medical data, where we will look at suggesting treatments based on the patients' medical history.

In real life it is very rare to find perfect similarity, thus we need to introduce a system of different levels of similarity. So then a statement such as *Bob is more like Charlie than Alice*, can be translated into partial similarity by saying that the similarity between Bob and Charlie is 0.8, whilst the

similarity of Alice and Bob is only 0.2. Therefore we can think of partial similarity as an extension of the similarity relations. In our case, we only used three values in the matrix of the similarity relation T: 1, -1 and 0. Then, for partial similarities we can extend this set of values to the interval $[-1, 1]$, therefore covering a range of different possibilities. The algorithm of von Mises can still be used in this case.

Previously, we have developed a method for describing the similarity of objects given by incomplete information [4], which can be traced back to partial similarity.

REFERENCES

[1] S Aeberhard, D Coomans, and O De Vel. Comparison of classifiers in high dimensional settings. *Dept. Math. Statist., James Cook Univ., North Queensland, Australia, Tech. Rep*, 92(02), 1992.

[2] Alon Altman and Moshe Tennenholtz. Ranking systems: the pagerank axioms. In *Proceedings of the 6th ACM conference on Electronic commerce*, pages 1–8. ACM, 2005.

[3] László Aszalós, Lajos Hajdu, and Attila Pethő. On a correlational clustering of integers. *Indagationes Mathematicae*, 27(1):173–191, 2016.

[4] László Aszalós and Tamás Mihálydeák. Rough classification in incomplete databases by correlation clustering. In *2015 Conference of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology (IFSA-EUSFLAT-15)*. Atlantis Press, 2015.

[5] R. V. Mises and H. Pollaczek-Geiringer. Praktische verfahren der gleichungsauflösung . *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik*, 9(2):152–164, 1929.

[6] Z. Néda, R. Sumi, M. Ercsey-Ravasz, M. Varga, B. Molnár, and Gy. Cseh. Correlation clustering on networks. *Journal of Physics A: Mathematical and Theoretical*, 42(34):345003, 2009.

Communication Papers of the Federated Conference on Computer Science and Information Systems pp. 21–28
DOI: 10.15439/2019F93
ISSN 2300-5963 ACSIS, Vol. 20

Correlation Clustering: Let All The Flowers Bloom!

László Aszalós
Faculty of Informatics at University of Debrecen,
26 Kassai út H4028 Debrecen , Hungary
Email: aszalos.laszlo@inf.unideb.hu

Mária Bakó
Faculty of Economics at University of Debrecen,
138 Böszörményi str., H4032 Debrecen, Hungary
Email: bakom@unideb.hu

***Abstract*—Correlation clustering is a NP-hard problem, and for large signed graphs finding even just a good approximation of the optimal solution is a hard task. In this article we examine the effect of ranking the nodes and processing them in order of ranks. We demonstrate that based on the rate of positive edges in the graph it is worth using different optimisation methods. We show that all building blocks of our methods are necessary under certain circumstances.**

I. Introduction

ONE OF the typical tools of unsupervised learning is cluster analysis. Here, we wish to put similar elements into the same group, and put different elements into different groups. The clustering methods are usually based on the distance or density of the objects. There is an exception—correlation clustering—where we only have a similarity relation of the objects. This is a tolerance relation, i.e. reflexive and symmetric, but not necessarily transitive. The result of a clustering is a partition of objects, which can be understood as an equivalence relation, i.e. a reflexive, symmetric and transitive relation. In 1965 Zahn asked the following question [14]: *What is the closest equivalence relation for a given tolerance relation?*

In 2004 Bansal et al. have showed that this task is NP-hard and gave a non-optimal solution [6], which of course only produces an approximate solution. [6] also contains a generalization of the original problem, where the edges have weights, and instead of distances we want to optimize the sum of weights of *misplaced* edges. The generalized problem has many applications (mainly in image processing), and there are numerous approximation methods available in the literature, each based on a different approach: from minimal cut to linear programming [1], [8], [9], [10], [11], [13].

The clustering problem itself can be formulated as a combinatorial optimisation task, and therefore we can apply the usual methods: hill-climbing, scattered search, bee algorithm, harmony search, etc. In recent years, the authors have examined most of these [2], [5] and moreover have developed a new method (contraction) to solve the original clustering problem. Various implementations of this new method were born, on one hand to accelerate the methods, and on the other to achieve as close an approximation as possible.

Obviously, a general method cannot offer the best approximation in every particular case. Therefore, we can achieve a better approximation than the methods described above if we choose the most fitting method for each individual task. The contraction itself is a universal method, i.e. it can be used for all tolerance relations; however, implementations can be developed to fit the specialities of the given problem to be solved. As there are no widely accepted benchmark tolerance relations, we use random, signed graphs to test our methods. The random signed graphs can be easily identified by partial tolerance relations. (A positive edge denotes that the relation holds between the nodes of the edge; a negative edge denotes that the relation does not hold between the nodes of the edge; a missing edge means partiality, that is we have no information about similarity or dissimilarity.) Two types of random graphs are common. The Erdős-Rényi graph is a dense graph, where there is an edge between any pair of nodes with probability p. For any random *colouring* of the same graph—that is we assign positive or negative signs to the edges—the distance between the tolerance relation (which belongs to the signed graph) and the closest equivalence relation is almost the same due the symmetry of this kind of graphs. This means that the colouring does not affect the distance between the tolerance and the equivalence relations. Phase transformation of correlation clustering—determining the size of the biggest cluster in the nearest equivalence relation—can be solved relatively easily [12].

Barabási-Albert's scale-free—later abbreviated to BA-type—graph is a sparse graph, so the number of edges will be proportional to the number of vertices. Because there are only a few edges in the graph, it follows that usually there are only a few adjacent vertices, therefore the *colouring* of a given edge has a much stronger effect as before. Furthermore, these kind of graphs are not symmetric. Hence the distance of tolerance relations correspond to the random colourings of the same graph and the closest equivalence relation can vary within wide limits, as Fig. 1.3. in [3] demonstrates this. There is not yet a conjecture for the phase transformation problem for BA-type graphs. Measurement results for large graphs could possibly help to formalise this conjecture. This is the reason why we wish to develop a both efficient and accurate method for BA-type graphs, and determine which methods are the best for certain conditions.

The structure of the article is the following: after the introduction, in Section II we present our notations including the elements needed to build our methods. Next we delineate the overall results, and review our experiments. In Section IV we discuss the results and the roles of the building blocks of our methods. Finally we conclude our results and we give our

©2019, PTI

further plans.

II. System of notation

A. Tolerance and equivalence relation

Let a set of objects be denoted by V, and their similarity (tolerance) relation by T, where $T \subset V \times V$. The partition p is a function: $p : V \to \mathbb{N}$, where objects x and y are in the same cluster if and only if $p(x) = p(y)$. Accordingly we can interpret the cluster of x as $g_x = \{y \in V | p(x) = p(y)\}$. Of course, the partition p determines an equivalence relation *where elements are in a common cluster in* p.

By the distance between the tolerance relation T and the equivalence relation based on p we mean the number of cases—called as a *conflict*—in which exactly one of the two relations holds, i.e.

- xTy and $p(x) \neq p(y)$ or
- xTy does not hold, but $p(x) = p(y)$.

Our task in case of a given tolerance relation T is to determine the partition p for which the distance between T and p is minimal, or in other words the number of conflicts is minimal. The exponentially growing Bell numbers [7] give the number of partitions, so the exhaustive search—apart from some special cases—is not applicable, hence we can only get approximate solutions.

Although it is not feasible in reality, but the objects can be thought of as magnets, where the similar ones are attracted to each other, and the different ones toss each other. If we left these magnets alone, after some time they would form a state of equilibrium. Here the magnets construct groups. These groups can be considered as clusters and the whole collection of clusters as partitions. The steps of the magnets we want to model are:

- another magnet group is more attractive for a magnet than its own group, so it is moved to that group
- two magnet groups attract each other, so these groups are joined.

The first step is to be later referred to as *motion*, while the other as *merging*. The value of attraction between objects x and y—denoted by $a(x, y)$—will be:

- 1, if the relation holds between them, i.e. the objects are similar,
- -1, if the relation does not hold between them, i.e. the object are dissimilar,
- 0, if the relationship is not interpreted between the two objects.

The attraction $a : V \times V \to \mathbb{N}$ of objects can be generalized as the attraction between an object and a cluster, that is $a : V \times 2^V \to \mathbb{N}$ and as the attraction between two clusters, that is $a : 2^V \times 2^V \to \mathbb{N}$ which are defined as $a(x, g) = \sum_{y \in g} a(x, y)$ and $a(g, h) = \sum_{x \in g} a(x, h)$, where g, $h \subset V$.

The summing here corresponds to the superposition known from physics. We left to the reader to check that at the motion and merging steps, the distance between T and the partition that is changed by either of these two steps will only decrease. Therefore, if we can no longer reduce the number of conflicts using motion and merging, then the resulting partition is actually a local minimum, and we stop.

B. Optimization methods

It is easy to check that for $V = \{a, b, c\}$ if aTb and aTc holds but bTc does not, then more local minima exist. In this simple case the number of conflicts (the distance) at these local minima are the same. In case of thousands of objects, the situation becomes very complicated, and many of the local minima aren't global minima, moreover there are significant differences between distances at minima.

The question is, in what order do we take our steps to avoid these traps, and get close to the global minima? [3] presented three structures, where the accuracy of the last one was up to twelve percent worse than the others'. So we omit this structure from this article, and so it is not shown in Fig. 1. First, we reiterate the other two structures. Then, we look at the different implementations of the merging and motions steps, and finally look at how these come together into methods.

In the case of the upper structure in Fig. 1—called *sequential*—a motion step is followed by a merging step, and if at least one of them can be executed, then the partition changes, and so we can try to reduce the conflicts again. If the partition has not changed during the motion and merging steps, our algorithm stops. In case of lower structure—called *repeated*—we apply the motion step until it reduces the distance between T and p. Once the motion step can no longer change the partition, we begin to merge. We repeat the merging step until it reduces the distance between T and p. If at least one merging step has made a reduction to the distance (denoted in the Fig. 1 by `any?`) in this cycle, that is, if the partition has been changed by some merging steps, then we can begin a new round, and test the motion step again. When the motion and merging steps do not change the partition, we must stop.

Whilst the motion step only changes the partition locally, and in small increments; the merging step—especially in the final stages of the optimisation, when we have big clusters—can cause significant global changes. This is the reason why the third structure—which is similar to the *repeated* structure, with the position of the motion and merging steps swapped—results in larger distances in the tests because the method reached a local minimum much earlier.

C. Alternative implementations of the motion step

To determine the motion step, the forces of $a(x, g)$ should be calculated for each object x and each cluster g, and for each object x we need to find its most attractive cluster h for which $a(x, h) = \max_g a(x, g)$. If the object x is moved to another cluster h—because that cluster is more attractive for the object than its own cluster, i.e. $a(x, h) > a(x, g_x)$—object x may become a status quo breaker there, and an earlier element of this new cluster is transferred to a third cluster and this chain reaction could continue. However, the previously calculated values of attractiveness of objects and clusters will be invalidated by changing the partition and need to recalculated.

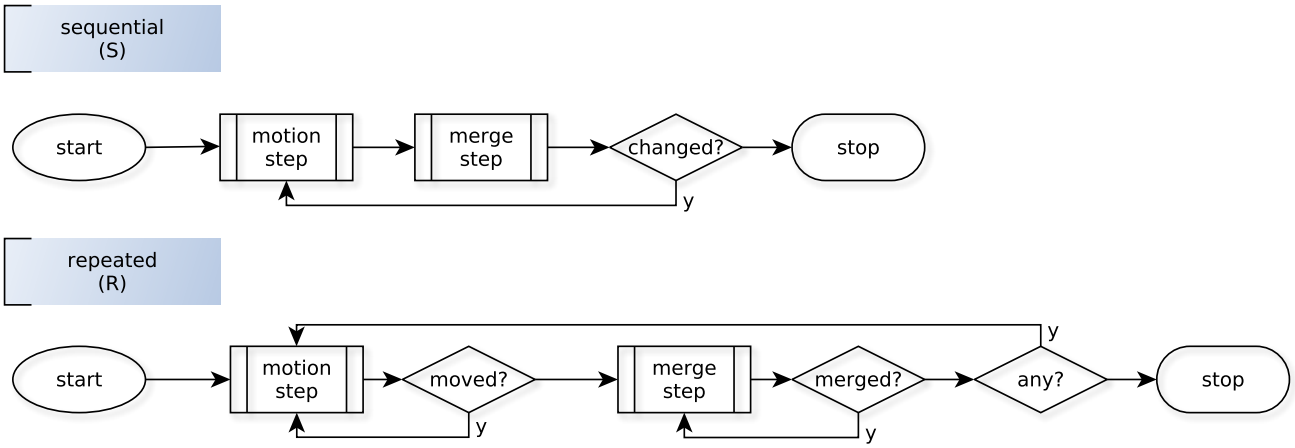


Fig. 1. Structure of our optimization methods.

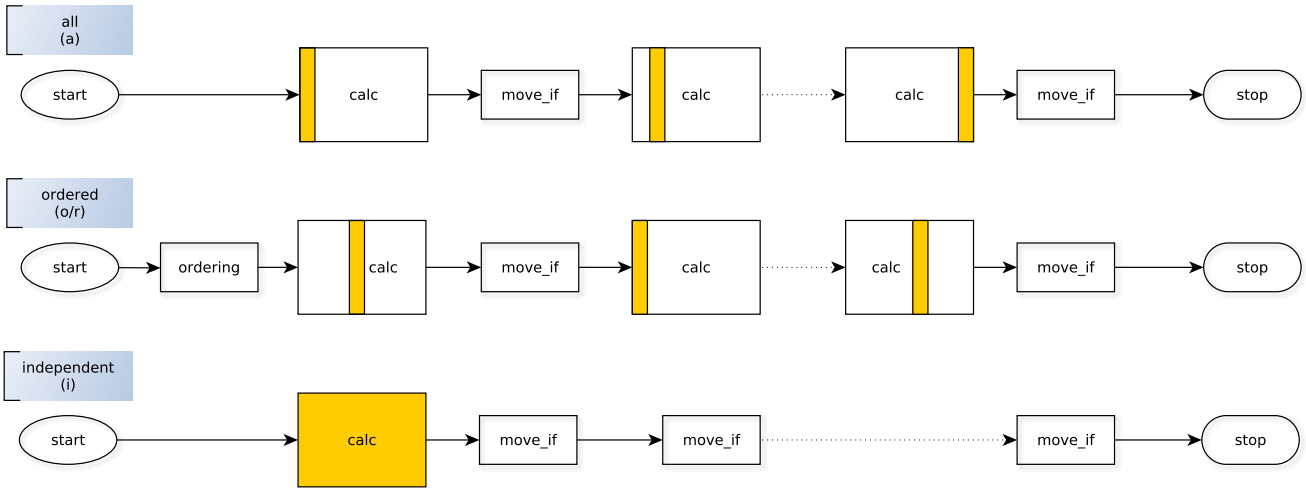


Fig. 2. Alternative flows of motion step.

We had an implementation of the motion step (flow *all* in Fig. 2) that takes all the objects one-by-one, and calculates the actual attractions of the objects. If some other (maybe empty) cluster is more attractive than its own cluster, it moves the object into that cluster. In the figure the colouring of the portion of the rectangle represents the range of the calculation. Here, we take every object $x \in V$ and calculate the values $\max_g a(x, g)$ and $a(x, g_x)$. The latter gives the attraction of its own cluster. If the former is larger - that is, another cluster has a greater attraction, then object x must move. Because we examine each object in a row, this process received the name *all*, and the abbreviation a. In the figure, the colouring of the corresponding rectangle shows that you do not need to compute attractions $a(y, g)$ for all objects y and for all clusters g, but only the ones that belong to object x: $a(x, g)$ for all clusters g. This determines the movement of x, and so we can move to the next object in the queue, and determine the attractions belonging to it.

The flow *independent* in Fig. 2—as the colouring shows—starts by calculating all the attractions $a(x, g)$ for all $x \in V$ and all clusters g. (That is why the entire rectangle is painted.) Next it determines the set

$$M = \{x \in V \,|\, \max_g a(x, g) > a(x, g_x)\},$$

i.e. the set of objects to move. In each step we need to select some object from the set M. If this object x moves from cluster g to the cluster h, then the set M in the next step will be reduced to $M \backslash (g \cup h)$ because the clusters g and h have changed, therefore the values previously calculated for these clusters are no longer valid and therefore should not be used. As the set M becomes empty, the process stops. Since the clusters affected by the motions (both from and to) are all different due to the deletions, the flow has been assigned the name *independent* and the abbreviation i.

We began to research the representatives of the clusters, and the methods to specify them [4]. To find the representatives of the clusters, a (initially unit) vector has to be multiplied by the matrix of the similarity relation and normalised again and again until a fixed point is obtained. In order to keep the running time under control, only a fixed number of steps were performed and the objects were ranked according to the result vector. Considering this rank, we can create two variants of the flow *all*, where this rank determines the order of processing, and not the (arbitrary chosen) identifiers of the objects. If we process objects in increasing order, we get the flow *ordered*, abbreviated by o, while if we choose the descending order, then flow *reversed*, or shortly r.

D. Alternative implementations of the merging step

For the merging step we use the attractions between clusters: $a(g, h)$. As merging causes a global change, a cautious approach allows merging two clusters only once. We are using the greedy approach, so we are interested in the most profitable merging, i.e. merge those two clusters for which the number of conflicts decreases the most. This is illustrated by the flow *one* in Fig. 3, and we will refer to it by O.

It is clear that the attractions of the merging clusters towards a third cluster will add up, so the attractions of the clusters can be maintained relatively easily. The flow *recalculate* (abbreviated by R) visualises this process. It is important to note—and this is also indicated by the colouring of the individual rectangles—that there is no need to recalculate all the attractions, therefore we can save ourselves from performing a lot of calculations by using this variant, instead of repeating flow *one*. Although the final result is exactly the same, the amount of calculation needed in the second case is only a fraction of the original.

The recalculation of forces may be omitted if we take the value of attractions in decreasing order, and we merge the clusters only if they have not yet changed in this process. This variant is called *independent*, and abbreviated by I.

III. Test results

A. Overall results

The combination of the structures and the implementations of the motion and merging steps provide an applicable method. In the names of the methods the first letter is the code of the structure (Fig. 1). The middle letter of the name of a method is the code of the motion step (Fig. 2), and the last letter is the code of the merging step (Fig. 3). Similarly to [3] we created a graph showing the results achieved with different methods.

We tested the various methods with 100 different BA type randomly generated tolerance relations of 5000 objects. We used the same tolerance relations as input for each method, and summed up the resulting conflicts and the running times of our Python implementations. These sums are presented in Fig. 4. Note that, the axis y has a logarithmic scale in order to equally allow for the demonstration of the fast and the slow methods.

The triangles in the upper left corner denote methods using the single merger (O). This location means that these are very accurate methods, but they can take up to ten times more time to run than the other two kinds of merging steps. In the lower right corner of the figure, crosses denote the methods applying the independent merger (I), these are fast but inaccurate.

We can find a coterie on the left side of Fig. 4. From left to right the pairs SaR-RaR, SrR-RrR and SoR-RoR overlap. While method RiR is above these (red circle), its pair, the SiR (black circle) is on the right side of the figure.

Based on this graph, we could think that the two new motion steps (r and o) are worse than the previous ones (a and i). Keep in mind that this figure was created by summing up the number of conflicts and running time, which can easily cover important details.

B. Detailed results

Let us consider a more sophisticated approach. We took 1200 different BA-type random tolerance relation of 500 objects, and examined 101 different colourings for each of them. Here the rate q of positive edges varies from 0 to 1, and we took the average of conflicts for each ratio.

From the slow methods, only SaO was kept in this test and it is represented with a triangle in the top left corner of Fig. 4, because it minimised the number of conflicts and therefore we consider it is as a benchmark. Since we approached the same distances with each of the 17 optimization methods, we would not really see any striking differences if we were to present the result on a typical graph. For this reason, we divide all the values by the corresponding values of the benchmark (Fig. 5). We can separate some of the methods, but most of them are very close to each other. However, it can be seen that some lines cross the level of 1, giving better results than the benchmark. We have no method that is always better than the benchmark, but we can find ones that are better for a small q, and ones are better for a big q.

For example method SaR can only tighten the benchmark around $q = 1$. The only difference between the methods SaR and SaO is that the latter merges once before attempting to move all the elements again, while the former continues to merge as long as possible. If only a few clusters with large number of elements constitute a partition, then there is no problem with successive merging, but in other cases this step causes too much change that cannot be counterbalanced by motion steps.

Consider the method RiR, which is dominated by the above-mentioned method SaR in the Pareto sense in Fig. 4, but rates of Fig. 5 shows that in many cases it exceeds the SaO method for both small and large q values. Therefore the Pareto set of Fig. 4 is not enough in itself to provide a method for finding the closest partition to a given tolerance relation.

Instead of presenting additional individual graphs, we made two summary figures (Fig. 6 and 7), where $q = 0$ corresponds to the left of the picture and $q = 1$ to the right. The methods with worst performance were not included in these figures. Fig. 6 shows the cases (method and q) which overtake the

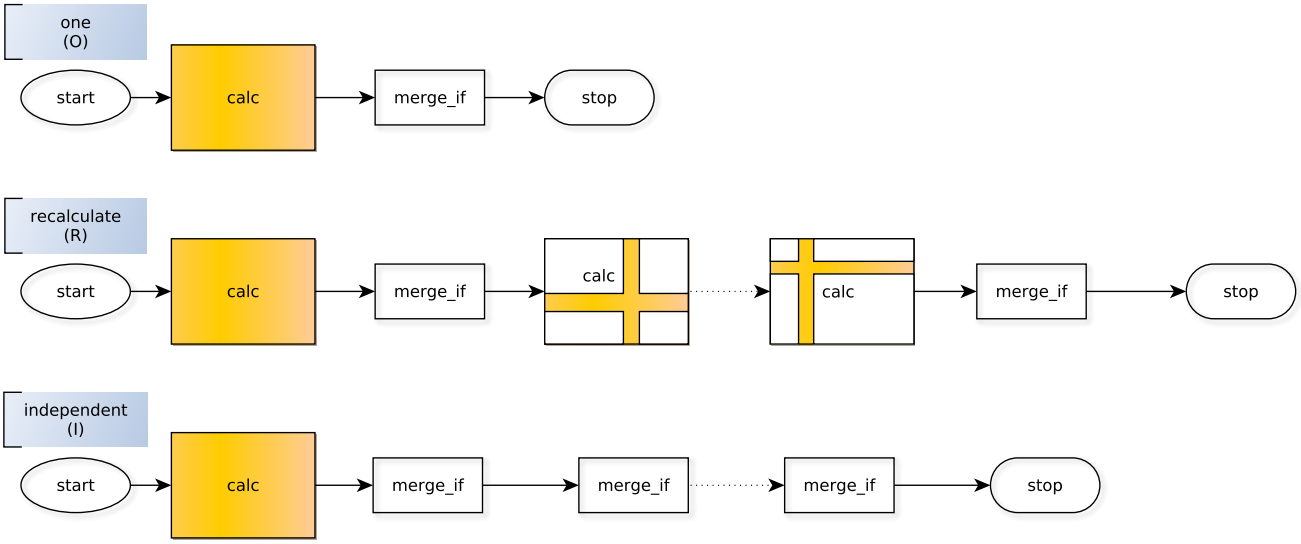


Fig. 3. Alternative implementations of merging step

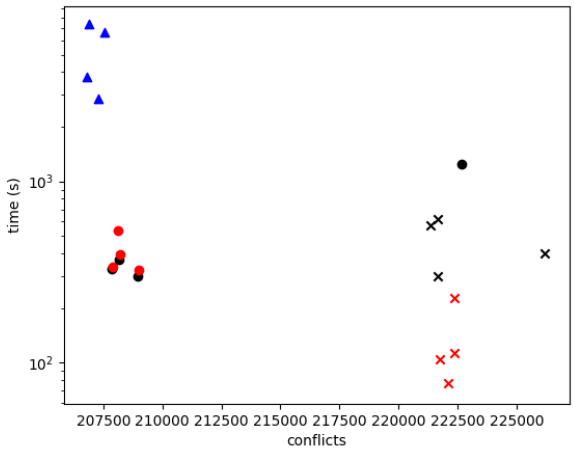


Fig. 4. Cumulative results on random relations: running time and efficiency of the variants of contraction. Shape and colour denote the merging step and the optimization method

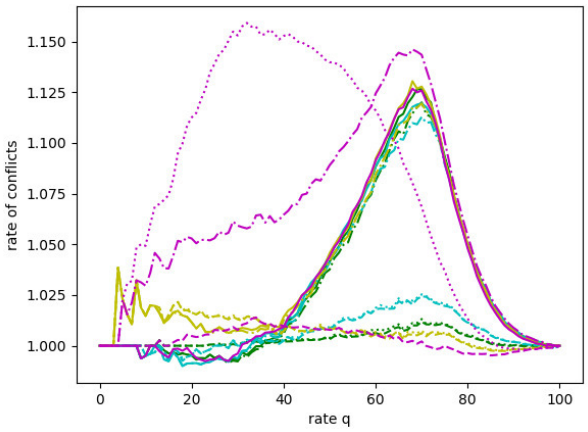


Fig. 5. Rate of number of conflicts and the benchmark value at different rates q of positive edges. Lower value is better.

benchmark, while Fig. 7 shows the cases when the benchmark is the winner. The greater the difference, the darker the colour. In order to make the differences more visible, in Fig. 7 we set the same black colour for the bigger differences.

It is interesting to compare the numbers and the coloured strips belonging to the methods; but for most of us these pictures and numbers contain too much information. So it follows that they should be compressed somehow, highlighting only the relevant information. For this reason, we have summed up the numbers (rates) grounding Fig. 5. We did not do a complete summing on interval $[0, 1]$—similarly to Fig.4—, but we constructed four subintervals and summed on them, as a

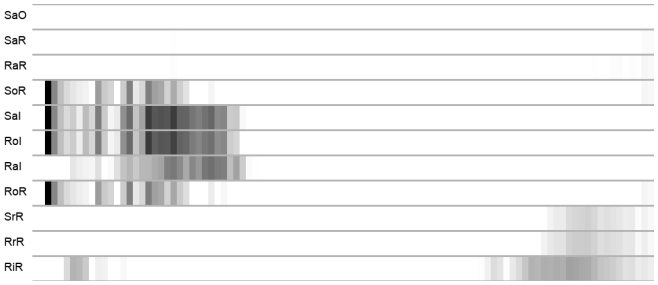


Fig. 6. Methods and rates q which give better result than the benchmark method SaO.

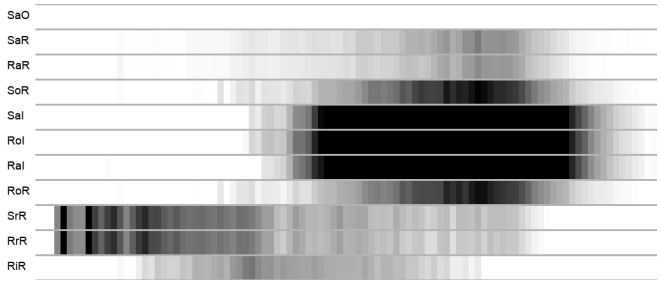


Fig. 7. Methods and rates q which give worse result than the benchmark method SaO.

TABLE I
SUBSUMS OF OUR OPTIMIZATION METHODS

Method	$\sum_{0}^{0.38}$	$\sum_{0.39}^{0.58}$	$\sum_{0.59}^{0.7}$	$\sum_{0.71}^{1}$
SaO	0.0	0.0	0.0	0.0
RaR	1.9	7.1	9.6	10.0
RiR	16.6	16.6	4.2	-7.2
SaR	2.3	8.1	11.3	11.1
RoR	-5.8	16.9	23.9	24.0
SoR	-5.2	18.2	25.0	24.8
RrR	54.8	14.7	6.1	1.4
SrR	55.7	15.7	7.3	2.6
SoI	-15.4	72.7	116.7	108.6
SaI	-10.9	68.3	120.3	111.0
RoI	-14.3	79.5	125.3	110.4
RaI	-9.5	75.3	129.2	112.2
RiI	-7.2	81.0	132.5	107.7
SrI	42.6	75.3	122.6	109.8
RrI	44.5	81.9	133.4	111.2
SiI	162.8	172.4	164.8	122.7
SiR	368.1	282.9	115.4	36.2

Table I shows. Here, the rows of the table are sorted by their sums, so the better methods are at the top, and the mostly weak methods are at the end in the table. Here the running time could not be taken into account because it was not recorded.

There are several negative numbers in the first column of Table I, i.e. several methods perform better for small q than the benchmark. As q is small, tolerance is achieved between a few pairs of objects, so many small clusters will be included in the resulting partition. Therefore the merging steps do not cause very significant changes, so we can iterate the merging of clusters inside one merging step. According to the information summarised in this table, the best results belong to the methods SoI and RoI. To verify our findings, we made some experiments similar to Fig. 4. However, whilst q was arbitrary in that random tolerance relations used, here we tighten it for the given intervals. Fig. 8 reinforces that these methods really are the best, where RoI is more than three times faster than SoI. Only method RaI precedes both of them in speed but not in precision, and these three methods are the Pareto set (that is, there does not exist a method which is faster and more accurate at the same time).

There are no negative numbers in the second and third columns of Table I, i.e. no method can overtake the slowly and

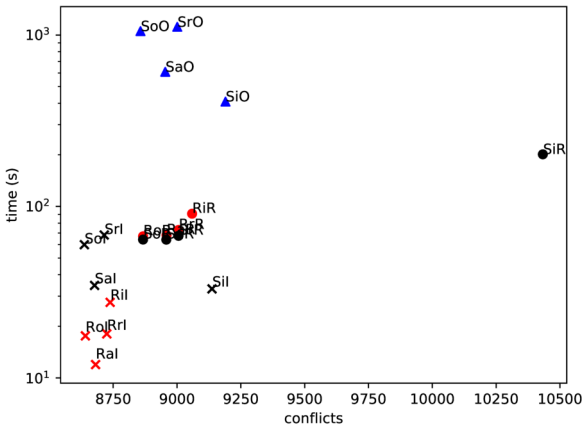


Fig. 8. Cumulative results on random relations where $q \in [0, 0.38]$: running time and efficiency of the variants of contraction

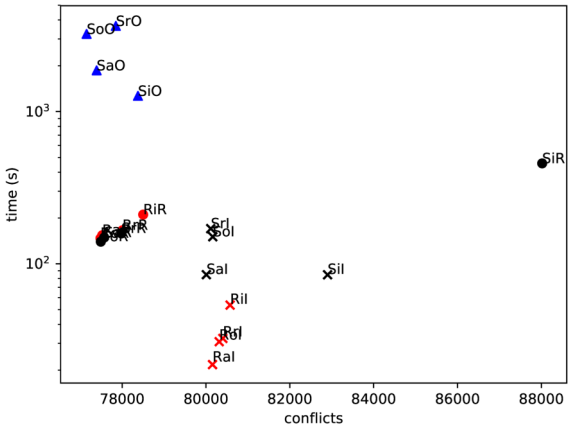


Fig. 9. Cumulative results on random relations where $q \in [0.38, 0.58]$: running time and efficiency of the variants of contraction

confidently evolving benchmark. Fig. 9 shows that the method SoO—not included in the table—is the absolute winner, even ahead of the benchmark. But it takes a long time to complete it. The method SoO is 22 times slower than the third ranked method RoR and 23 times slower than fourth ranked method SoR.

The methods RaR-SaR, which got small numbers in the second column of the table are the fifth and sixth most effective methods, and the methods SrR-RrR have scored the eighth and ninth place, but the methods RoR-SoR are better than all of them in the Pareto sense. The method SiO marked with the fourth triangle is the tenth, and there are some more efficient (faster and more precious) methods.

In Fig. 10—the figure representing the third column—triangles are located in the first five places, only the method RiR slipped in between them in the fourth place. That is, the principle *haste makes waste* really applies here. Looking

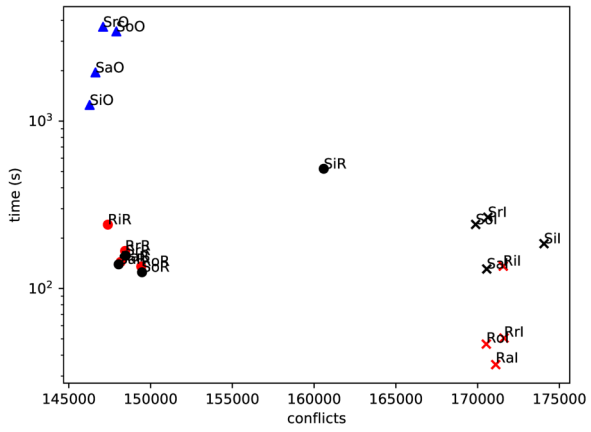


Fig. 10. Cumulative results on random relations where $q \in [0.59, 0.7]$: running time and efficiency of the variants of contraction

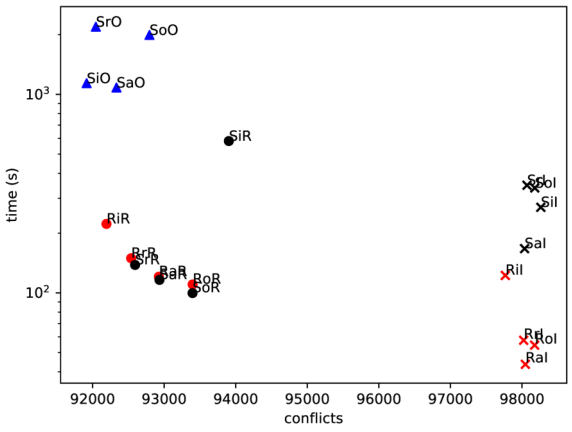


Fig. 11. Cumulative results on random relations where $q \in [0.71, 1]$: running time and efficiency of the variants of contraction

at the captions of the figure, we can see that most of the conflicts are here, so we have to handle the merging steps very carefully. The greedy independent merger performs poorly, all the crosses are on the right.

In the last column of Table I there is only one negative number and it belongs to the method RiR. The numbers belonging to methods RrR and SrR are small, but positive. If we narrow the last interval to $[-0.81.1]$, then all three numbers would turn negative (while the others would remain be positive), and their values would be -5.4, -2.8 and -2.2 respectively. These are the three methods in the last three rows of Fig. 6, and only these strips have significant non-white parts on the right. The method RiR is overcome by the methods SiO and SrO, but their runtime is nearly 5 and 10 times higher. On the other hand, RiR defeated our benchmark in the last interval, and the benchmark followed by the aforementioned methods RiR and SiR.

IV. Discussion

When wishing to use these optimisation methods in practice, it is possible to choose a precise or fast method depending on the desired outcome, along with the q value of the tolerance relation. The Pareto set is the set of methods that are optimal in that sense. That is, there is no such method which is better in both speed and precision than those in the Pareto set, so it is worth choosing a method contained in this set. Table II shows for each of the intervals which methods are optimal in this sense, and what are their ranks among the examined methods in accuracy and speed. For example, the method RaI in the first row of the table is the fastest of all, but if we disregard the first interval, then it performs with very poor accuracy.

Consider the building blocks of our methods and look at their roles. The *sequential* structure provides the most accurate solutions in each case, finding the closest equivalence relation for the tolerance relation of the problem. However, this often takes a lot of time. For a few tasks, a half-percent error can be sacrificed in accuracy for a fractional time calculation. That is why we may need a *repeated* structure.

TABLE II
Pareto Sets on different intervals

Method	$[0, 0.38]$	$[0.39, 0.58]$	$[0.59, 0.70]$	$[0.71, 1]$
RaI	4/1	14/1	18/1	16/1
RaR				8/7
RoI	2/2		14/2	
RoR		3/8	10/6	
RrI				14/3
RrR				5/10
RiI				
RiR			4/13	3/12
SaI		12/6		
SaR			6/8	9/6
SoI	1/7			
SoR		4/7	11/4	10/4
SrI				
SrR				6/9
SiI				
SiR				
SaO		2/18		
SoO		1/19		
SrO				
SiO			1/17	1/18

The implementation *recalculate* of the merging step can be interpreted as a compromise between the two structures, as this implementation is practically a cycle around the interpretation *one*. The interpretation *recalculate* is indicated by a circle in our figures. Except for the first interval, these circles are shown below the triangles (*one*) and to left on the crosses (*independent*), so they can also be considered as a compromise in this sense, they try to balance between the speed and the accuracy.

Apart from the first interval, the triangles (*one*) occupy the left upper part of the figures, resulting in precise solutions and

even the most accurate ones.

The crosses only perform well in the first interval, but then there are dominant.

When we turn to the implementation of motions, the implementation of *ordered* performed very nicely as a new player. The fact that we start by first processing the low-ranking objects—that cause the small changes—has clearly proven to be efficient in the first two intervals, reaching two-two podium positions. If we compare the efficiency in the lower left corner of the figure (ignoring the speed) we get the order of o-a-r-i, i.e. by fixing the structure and merging step we go from the most powerful method to the weakest.

In the third and fourth intervals, we cope with orders i-a-r-o and i-r-a-o, respectively. Partly this is the reason that in the third interval the method SaR is also a Pareto optimum beside the method RiR, and the method of RaR is just behind SaR. In the fourth interval, for similar reasons we can discover the Pareto optimum chain of methods RiR-RrR-SrR-RaR-SaR-SoR. Therefore, by prioritising the speed of the implementations *all* and *reversed* can play a role.

The motion step's *independent* implementation is very divisible. The SiI and SiR methods perform poorly for every interval. However, RiR is considered to be the Pareto optimum in the last two intervals, and it is very close to it in the second interval too. Additionally, the most accurate method of the last two intervals is SiO. The latter may be due to the fact that larger clusters are formed here, so moving an object from cluster g to cluster h in this implementation means that elements of g and h will be excluded from moving in that step. So from the four moving methods, it moves the least amount of items. Implementation O also means a one-time merging, so we change our partition the least as possible with this method. Many of these small changes—together—give the closest partition.

If we summarise the above statements, we see that each implementation and structure has its own place, and it is worth using different methods under different circumstances (for different q values).

V. Conclusion and further work

We've reviewed our previously described methods and implemented two new ways to move our objects in order of their importance and process them accordingly. We carried out a much larger and more thorough experiment than before, which included the new methods too. We elected four intervals of $[0, 1]$ in which the methods perform differently. By restricting our experiments to these intervals we determined which methods should be used for a given tolerance relation.

Python implementations were suitable for prototyping, testing new and newer ideas. Now we wish to develop a faster (probably C++) implementation that constantly updates the values of forces $a(x, g)$ when changing the partition, and stores whether the information on objects that need to be moved or joined is valid or outdated.

This can eliminate unnecessary calculations, which can speed up our methods. It should be taken into account how much energy such an administration needs, and whether it makes the software faster or slower.

References

[1] Agarwal, G., Kempe, D.: Modularity-maximizing graph communities via mathematical programming. The European Physical Journal B **66**(3), 409–418 (2008)

[2] Aszalós, L., Bakó, M.: Advanced search methods (in Hungarian). http://www.tankonyvtar.hu/hu/tartalom/tamop412A/2011-0103_13_fejlett_keresoalgoritmusok (2012)

[3] Aszalós, L., Bakó, M.: Contraction methods for correlation clustering: The order is important. In: Recent Advances in Computational Optimization, pp. 1–13. Springer (2019)

[4] Aszalós, L., Nagy, D.: Iterative set approximations based on tolerance relation. In: International Joint Conference on Rough Sets, pp. 76–88. Springer (2019)

[5] Bakó, M., Aszalós, L.: Combinatorial optimization methods for correlation clustering. In: D. Dumitrescu, R.I. Lung, L. Cremene (eds.) Coping with complexity, pp. 2–12. Casa Cartii de Stiinta, Cluj-Napoca (2011)

[6] Bansal, N., Blum, A., Chawla, S.: Correlation clustering. Machine Learning **56**(1-3), 89–113 (2004). DOI 10.1023/B:MACH.0000033116.57574.95. URL http://dx.doi.org/10.1023/B:MACH.0000033116.57574.95

[7] Berend, D., Tassa, T.: Improved bounds on bell numbers and on moments of sums of random variables. Probability and Mathematical Statistics **30**(2), 185–205 (2010)

[8] Bonchi, F., Gionis, A., Gullo, F., Tsourakakis, C.E., Ukkonen, A.: Chromatic correlation clustering. ACM Transactions on Knowledge Discovery from Data (TKDD) **9**(4), 34 (2015)

[9] Demaine, E.D., Emanuel, D., Fiat, A., Immorlica, N.: Correlation clustering in general weighted graphs. Theoretical Computer Science **361**(2-3), 172–187 (2006)

[10] Emanuel, D., Fiat, A.: Correlation clustering–minimizing disagreements on arbitrary weighted graphs. In: European Symposium on Algorithms, pp. 208–220. Springer (2003)

[11] Giotis, I., Guruswami, V.: Correlation clustering with a fixed number of clusters. In: Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm, pp. 1167–1176. Society for Industrial and Applied Mathematics (2006)

[12] Néda, Z., Florian, R., Ravasz, M., Libál, A., Györgyi, G.: Phase transition in an optimal clusterization model. Physica A: Statistical Mechanics and its Applications **362**(2), 357–368 (2006). DOI 10.1016/j.physa.2005.08.008. URL http://dx.doi.org/10.1016/j.physa.2005.08.008

[13] Shi, J., Malik, J.: Normalized cuts and image segmentation. Departmental Papers (CIS) p. 107 (2000)

[14] Zahn Jr, C.: Approximating symmetric relations by equivalence relations. Journal of the Society for Industrial & Applied Mathematics **12**(4), 840–847 (1964). DOI 10.1137/0112071. URL http://dx.doi.org/10.1137/0112071

Communication Papers of the Federated Conference on Computer Science and Information Systems pp. 29–36
DOI: 10.15439/2019F85
ISSN 2300-5963 ACSIS, Vol. 20

Exact and approximation algorithms for joint routing and flow rate optimization

Konstanty Junosza-Szaniawski
Warsaw University of Technology
Faculty of Mathematics and Information Science
ul. Koszykowa 75, 00-662 Warszawa, Poland
Email: k.szaniawski@mini.pw.edu.pl

Dariusz Nogalski
Military Communication Institute
C4I Systems Department
ul. Warszawska 22A, 05-130 Zegrze, Poland
Email: d.nogalski@wil.waw.pl

***Abstract*—This paper addresses the comparison of algorithms for a version of the Network Utility Maximization (NUM) problem. The joint formulation of routing and transmission rate control within the multi-user and single-path setting is assumed within the NUM. Since the problem is NP-hard, the efficient heuristics are designed, implemented and compared experimentally with other existing heuristics and exact linear programming solver. The linear approximation is applied for a nonlinear utility function. The results of the experiments demonstrate a trade-off between computing time and precision of goal value.**

***Index Terms*—Mathematical programming, network optimization problem, network utility maximization, rate allocation, congestion control, multi-rate computer networks, routing, NP-hard problem, heuristic algorithm**

I. Introduction

The Network Utility Maximization (NUM) problem was introduced by Kelly et al. in 1998 [10]. Within the problem a single route is associated with a user (constant to the problem) and the goal is to maximize aggregate utility of all user rates (variable to the problem). Such a basic NUM problem is known to be polynomial-time solvable [17]. We refer to such a problem as the *basic NUM*.

This article addresses an extended problem where both user rates and routes are variable to the problem. We refer to such a problem as the *extended NUM* problem or simply, the *NUM* problem. Wang et al. prove such a problem to be *NP-hard* [17]. As a consequence, efficient heuristics are needed for larger networks.

Such a joint formulation of the problem is useful since path selection determines link congestion and the problem can be solved *efficiently*. Efficiency is becoming a critical problem in many application areas, e.g. in multimedia WSN (Wireless Sensor Networks). Multimedia data producing wireless sensor nodes require higher bandwidth, higher energy usage, and strict quality of service (QoS) requirements [13]. Additionally, we are interested in techniques to achieve *fairness* of multi-stream allocations [10] [11] [3].

We designed and implemented several heuristics and compared their performance with other existing heuristics and the exact linear programming solver. Due to the fact that our utility function was nonlinear, we applied linear approximation within optimization procedure. The results of the experiments demonstrate a trade-off between computing time and precision of algorithms.

Since our approach is centralized, it cannot be used directly in multi-rate networks operating in a distributed way. However, there are certain applications where our heuristics can be found useful. Firstly, in applications where centralized network controllers are applied, e.g. SDN. Secondly, to estimate a gap between single and multiple path solutions [18].

II. Motivation and related work

A. The complexity of various versions of network utility maximization problem

The complexity of the basic and the extended NUM was stated in the introduction. However, for various problem modifications it may differ. The decision version of m-commodity flow problem with fixed rates and single-path setting is NP-complete [4]. It is proved by reduction from the decision version of the bin-packing problem. In [5] it is shown that even the decision version of two-commodity integral flow problem is NP-complete, by reduction from the SAT problem.

B. Distributed models

Since the work of Kelly et al. [10] was published, there have been many works on price-based distributed network utility maximizing methods. Chiang et al. [3] present the TCP/IP joint optimization example and solve it through distributed decomposition. Horizontal decomposition of TCP-AQM Congestion Control to share link capacities among competing users is described. Vertical decomposition of TCP/IP Joint Congestion Control and Routing is described.

C. Centralized control in SDN networks

Amokrane el al. [1] propose an online energy efficient flow-based routing approach for SDN controller, which allows for dynamic reconfiguration of existing flows (paths and rates). Huang et al. [8] solve utility-optimized flow-level bandwidth allocation in hybrid SDNs.

D. Importance of single-path routing

Single-path routing is still important from the practical point of view in the following areas: hybrid SDN networks [8], energy-aware networks [1] [9], IP networks [3] [10],

©2019, PTI

MANETs (Mobile Ad hoc NETworks) [16] and WSNs [13]. It imposes a smaller overhead in implementation and is less expensive to support, as opposed to the multi-path [18]. Multi-path routing, as opposed to single-path routing has several disadvantages/challenges: data packet reordering, route maintenance, route discovery and a computational overhead.

E. Fairness and Efficiency

In literature there are various definitions of *fairness*. Two of them are most common: Max-min fairness [14], and proportional fairness [10] [14] [19] [11] [3]. The *proportional fair* vector comes as a solution to maximization of aggregate utility given by the equation (11) . Utility functions can be interpreted as the level of elasticity of application traffic, user satisfaction, optimality of resource allocation efficiency, as well as fairness [3]. If $\alpha = 0$, the NUM problem reduces to system throughput maximization. If $\alpha = 1$, proportional fairness among competing users is attained. And finally, if $\alpha = \infty$, then max-min fairness is achieved (Lee et al. [11]). The improvement in fairness is achieved by sacrificing the network efficiency. As the value of α increases, the total rate decreases. Our problem formulation assumes $\alpha = 1/2$ for heuristics testing purpose.

F. Multicommodity flow problem and its heuristics

Our problem belongs to a class of more general flow problems, namely the multicommodity flow problem (MFP) (Leighton et al. [12]). However, our formulation is single-path, as opposed to the multi-path formulation in MFP. Additionally, we do not assume a fixed demand for each source-destination pair, but express such a demand with utility function. MFP heuristics may be found in many works e.g. Garg and Konemann [7], Fleisher [6].

G. Heuristics for NUM with single-path routing

Drwal [4] defines Utility-Maximizing Network Design problem understood as joint optimization over rates and paths (single-path). The work introduces Maximum Spanning Tree based algorithm, which gives a solution not less than a factor $O(\frac{1}{m})$ of an optimal solution, where m is the number of pairs.

Wang et al. [18] introduce several upper-bounds to estimate the performance gap between multi- and single-path solutions of joint optimization of transmission rates and paths. They introduce two heuristics: vertex-projection and greedy-branch-and-bound.

III. Main results

In our paper we compared experimentally several heuristics to the NUM problem: MstNum (see alg 1), IterativeMstNum (see alg 2), LPRoundingNum (see alg 3), LPBestPathNum (see alg 4) and IterativeLPBestPathNum (see alg 5).

Our main contribution is the design of two heuristics: IterativeMstNum (an iterative procedure based on maximum spanning tree) and IterativeLPBestPathNum (an iterative procedure based on LP solver).

The other heuristics may be found in literature: MstNum [4], LPRoundingNum [4] [18] and LPBestPathNum [18].

All the algorithms are implemented and compared experimentally with MilpNum (see alg 6), which gives an optimal solution.

IV. Problem definition

Let's assume:

1) $G = (V, E)$ an undirected graph representing a network, where V is a set of nodes and E a set of edges (this model could also be defined for a directed graph)
2) $c : E \to [0, \infty)$ capacity function
3) S a set of users
4) $F_{st} = \{(s_k, t_k) \in V \times V : k \in S, \text{ and } s_k, t_k \in V\}$ a set of (source,target) pairs - demands for transmission
5) $u : S \times [0, \infty) \to [0, \infty)$ utility function
6) $u(k, x) = u_k(x)$ utility value of transmission rate x for the user $k \in S$

The problem is to find a set of pairs $\{(x_k, P_k) : k \in S\}$, where x_k is a non-negative transmission rate assigned to the user (2), P_k is a transmission path in the graph G from s_k to t_k, assignments do not exceed (altogether) the edge capacity (3) and rates are maximized (1). Such an extended NUM problem is a generalization of the basic NUM formulated by Kelly [10].

$$\max \sum_{k \in S} u_k(x_k) \tag{1}$$

For the purpose of the article we assume $\forall_{i,j,\in S} \;\; u_i = u_j$.

$$\forall_{k \in S} \;\; x_k \geq 0 \tag{2}$$

$$\forall_{\{u,v\} \in E(G)} \sum_{k, \text{ s.t. } \{u,v\} \in E(P_k)} x_k \leq c(u, v) \tag{3}$$

MILP formulation: In MILP each flow f_k takes a single path P_k from the source s_k to the target t_k. We introduce a real variable f_{kuv} which indicates the amount of flow f_k passing via edge (u, v). The variable f_{kuv} is defined to indicate $u \to v$ flow direction and f_{kvu} reverse. We rewrite the constraint (3) as (4). We set a requirement that the flow is balanced (5), except source and terminal nodes (6)(7).

$$\forall_{\{u,v\} \in E(G)} \sum_{k, \text{ s.t. } \{u,v\} \in E(P_k)} f_{kuv} + f_{kvu} \leq c(\{u, v\}) \tag{4}$$

$$\forall_{k \in S} \;\; \forall_{u \in V \setminus \{s_k, t_k\}} \sum_{v:\{v,u\} \in E(G)} f_{kvu} = \sum_{w:\{u,w\} \in E(G)} f_{kuw} \tag{5}$$

$$\forall_{k \in S} \sum_{w:\{s_k,w\} \in E(G)} f_{ks_kw} = x_k \tag{6}$$

$$\forall_{k \in S} \sum_{v:\{v,t_k\} \in E(G)} f_{kvt_k} = x_k \tag{7}$$

We introduce a binary variable y_{kuv} which indicates whether flow f_k is passing via edge (u, v) (8). The variable

y_{kuv} is defined to indicate $u \to v$ flow direction and y_{kvu} reverse.

$$\forall_{k \in S} \ \forall_{\{u,v\} \in E(G)} \ y_{kuv}, y_{kvu} \in \{0,1\} \tag{8}$$

We bind the variable y_{kuv} with the variable f_{kuv} using two additional constraints (9)(10).

$$\forall_{kuv:k \in S, \{u,v\} \in E(G)} \ f_{kuv} \leq y_{kuv} \cdot c(\{u,v\}) \tag{9}$$

$$\forall_{kuv:k \in S, \{u,v\} \in E(G)} \ y_{kuv} \leq f_{kuv} \cdot M \tag{10}$$

where M is a large number.

In practice, *fair* allocations of capacity are needed. This can be obtained by choosing the utility function properly. In literature [3] [10] [4] most common is a family of functions:

$$u_k(x_k) = \begin{cases} w_k \frac{1}{1-\alpha} x_k^{1-\alpha}, & \alpha > 0, \alpha \neq 1 \\ w_k log(x_k), & \alpha = 1 \end{cases} \tag{11}$$

For the experiment purpose, we have chosen $\alpha = 1/2$, which gives the function (12). We assume flow value to be non-negative (2).

$$\forall k \in S \ \ u_k(x) = u(x) = 2\sqrt{x} \tag{12}$$

We approximate the function (12) with three linear functions over range $x \in [0,1]$ (see fig 1). Such a choice stems from our assumption $\forall \{u,v\} \in E(G) \ \ c(u,v) = 1$. If links differ in terms of capacity, approximation function could range from 0 to max edge capacity. For the approximation purpose, an additional equation is introduced (14). The goal function (1) is rewritten as (13).

$$\max \sum_{k \in S} (a_1 \cdot t_{1,k} + a_2 \cdot t_{2,k} + a_3 \cdot t_{3,k}) \tag{13}$$

where $a_1 = 12.00, a_2 = 3.00, a_3 = 1.3(3)$

$$\forall_{k \in S} \ \ x_k = t_{1,k} + t_{2,k} + t_{3,k} \tag{14}$$

where $0 \leq t_{1,k} \leq 1/36$, $0 \leq t_{2,k} \leq 8/36$, and $0 \leq t_{3,k}$

Relaxed LP formulation: We relax the y_{kuv} variable to allow the flow to split and flow via multiple paths (15). Such relaxation formulates the problem known in literature as Multi-Commodity Flow (MCF). Since this is a linear program, MCF can be solved in polynomial time [15].

$$\forall_{k \in S} \forall_{\{u,v\} \in E(G)} \ y_{kuv}, y_{kvu} \in [0,1] \tag{15}$$

In this NUM formulation all constraints { (2), (4), (5), (6), (7), (14) } hold except {(8),(9),(10)}, which are replaced with (15).

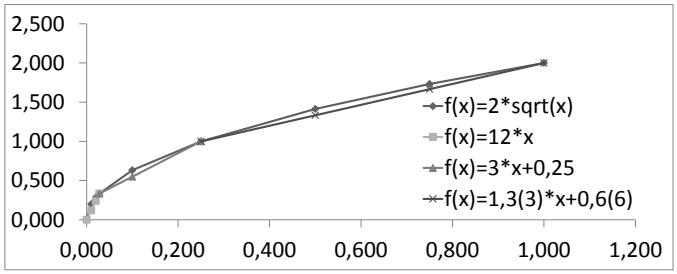


Fig. 1. Approximation of utility function $u_k(x) = 2\sqrt{x}$

V. ALGORITHMS

To solve the NUM problem we have implemented the algorithms below. Some of them have been designed by us and some taken from literature (see chapter III Main results).

1) *MstNum*(G, c, F_{st}) (see alg 1)
2) *IterativeMstNum*(G, c, F_{st}) (see alg 2)
3) *LPRoundingNum*(G, c, F_{st}) (see alg 3)
4) *LPBestPathNum*(G, c, F_{st}) (see alg 4)
5) *IterativeLPBestPathNum*(G, c, F_{st}) (see alg 5)
6) *MilpNum*(G, c, F_{st}) (see alg 6)

All the algorithms assume the following *input* parameters:

- G graph representing a network
- c capacity function
- F_{st} a set of demands for transmission

and *output* parameters:

- FAS such a set of pairs $\{(x_k, P_k) : k \in S\}$ that constraints (2)(3) are satisfied
- U aggregate utility function value.

A. MST based NUM

The algorithm *MstNum* is based on [4]. It uses single MST (Maximum Spanning Tree) run (line 1). All flow paths belong to the MST (line 3). If competing flows pass the same edge, the edge capacity is shared among the flows proportionally with the use of *min* function (line 11).

In the worst case (having a network with $|E| = m$ parallel edges spanned between two vertexes), assuming $\forall_{e \in E} c(e) = k$, and $|S| = m$, the heuristics *MstNum* gives a ratio of a solution to an optimal solution equal to $1/m$. It is shown in [4] that the algorithm never returns a solution that is less than a factor $O(\frac{1}{m})$ of an optimal solution.

B. Iterative MST based NUM

The algorithm *IterativeMstNum* offers a large improvement over *MstNum* with respect to maximization of the goal function. It consists of two phases (init phase and iterative phase). The first phase (init) runs *MstNum* (line 1). The second phase runs an iterative utility improvement procedure (lines 3-28). Within each iteration (a single *while* loop), $|S|$ solutions are found (line 4-21), compared and the best one is selected (line 22) to produce the input flow allocation set (FAS) for the next iteration (line 26). To find an i-th solution (within an iteration) the algorithm produces i-th MST (line 16) taking $c^{'}$ as the

Algorithm 1 MstNum(G, c, F_{st})

1: $T = MST(G, c, F_{st})$
2: **for** $i \in S$ **do**
3: Let P_i be $s_i t_i$-path in T
4: Update FAS with P_i
5: $x_i^{'} = min_{\{u,v\} \in E(P_i)} \ c(u,v)$
6: $x_i = x_i^{'}$
7: **end for**
8: **for** $e \in E(T)$ **do**
9: let $Shr_e = \{i \in S : e \in E(P_i)\}$
10: **if** $(|Shr_e| > 1)$ **then**
11: $\forall_{k \in Shr_e}$ let $x_k = min(x_k, \frac{x_k^{'}}{\sum_{\ell \in Shr_e} x_\ell^{'}} \cdot c(e))$
12: Update FAS with x_k
13: **end if**
14: **end for**
15: **return** $(FAS, \sum_{k \in S} u(x_k))$

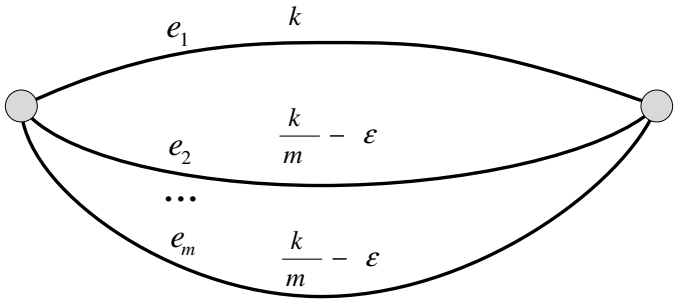


Fig. 2. Worst case for IterativeMstNum

graph edge capacity (lines 6-15), which equals c decreased by the capacity used by the other users, and re-routes i-th flow to a possibly better location within the i-th MST (the new route may be the same as the previous one) (line 17). The Δ_i is computed as the aggregated sum of i-th flow utility difference and all j-th flow utility difference (line 20). The best solution is selected Δ_{max} (line 22). If the objective improvement is less or equal to Δ_ε, the stop condition is satisfied (line 24).

In the worst case (see fig 2), assuming $|S| = m$, a ratio of a solution to an optimal solution is determined by the equation (16). For $m \gg 1$ the ratio approximates to $1/2$. The algorithm *IterativeMstNum* never returns a solution that is less than a factor $O(\frac{1}{2})$ of an optimal solution.

$$\frac{sol}{opt} = \frac{1}{2 - 1/m} \tag{16}$$

However, for practical networks the heuristics exposes a solid ratio (see chapter VI Computational results).

C. Linear program relaxation and rounding NUM

The algorithm *LPRoundingNum* (a slight improvement of the one given in [4]) begins with solving linear program relaxation in phase1 (line 3). Subsequently in phase2 (lines 5-14) a rounding is done by random selection of the single path for each user. Finally, in phase3 (line 16) the LP problem is solved again taking fixed single paths identified in the previous phase as a constraint.

Algorithm 2 IterativeMstNum(G, c, F_{st})

1: $FAS = MstNum(G, c, F_{st})$
2: $stop = false$
3: **while** $(stop == false)$ **do**
4: **for** $i \in S$ **do**
5: $(x_i^{'}, P_i^{'}) = FAS(i)$
6: **for** $(u,v) \in V \times V$ **do**
7: $c^{others}(u,v) = 0$
8: **for** $j \in S$, s.t. $j \neq i$ **do**
9: $(x_j^{'}, P_j^{'}) = FAS(j)$
10: **if** $(u,v) \in P_j^{'}$ **then**
11: $c^{others}(u,v) = c^{others}(u,v) + x_j^{'}$
12: **end if**
13: **end for**
14: $c^{'}(u,v) = c(u,v) - c^{others}(u,v)$
15: **end for**
16: $T_i = MST(G, c^{'}, F_{st})$
17: Let $P_i^{''}$ bet $s_i t_i$-path in T_i
18: $x_i^{''} = min_{(u,v) \in P_i^{''}} \ c^{'}(u,v)$
19: $\forall_{j \in S, j \neq i}$ recalculate $x_j^{''}$ using the idea given in alg 1 (lines 2-14)
20: Compute $\Delta_i = u(x_i^{''}) - u(x_i^{'}) + \sum_{j \in S, j \neq i}(u(x_j^{''}) - u(x_j^{'}))$ using (12)
21: **end for**
22: $\Delta_{max} = max\{\Delta_i\}_{i \in S}$
23: **if** $(\Delta_{max} \leq \Delta_\varepsilon)$ **then**
24: $stop = true$
25: **else**
26: Update FAS with the best solution in the current iteration $(x_{max}^{''}, P_{max}^{''})$
27: **end if**
28: **end while**
29: **return** $(FAS, \sum_{k \in S} u(x_k))$

D. LP Best Path NUM

The algorithm *LPBestPathNum* begins with solving linear program relaxation (line 3). Within the relaxed LP solution, for each $k \in S$ the single max path (max min y_{kuv} on the path) is identified as P_k (line 6). The procedure is using the BFS (Breadth First Search) for that purpose. Finally, the LP problem is solved again taking fixed single paths identified (line 9) as a constraint. The algorithm is similar to the vertex projection method given in [18].

E. Iterative LP Best Path NUM

The algorithm *IterativeLPBestPathNum* constitutes $|S| + 1$ iterations. In each $\{1, .., |S|\}$ iteration, linear program relaxation is solved (line 7). Within the relaxed LP solution, for each $i \in S \setminus S'$ the single max path P_i (max min y_{iuv} on the path) is selected (line 10). Among the $|S \setminus S'|$-paths, the best (max) P_i is selected and added to the model (line 14) as

Algorithm 3 LPRoundingNum(G, c, F_{st})

1: Create a *problem* with goal as (13)
2: Add constraints {(2),(4),(5),(6),(7),(14),(15)} to the $problem$
3: Solve the $problem$
4: Retrieve $\{y_{(k,u,v)\in S\times V\times V}\}$ and U from the $problem$ solution
5: **for** $k \in S$ **do**
6: Set vertex $u \in V$ s.t. $u = s_k$
7: **while** $(u! = t_k)$ **do**
8: Let's define a set $R_u = \{v_i\text{: } (u, v_i) \in V \times V \text{ and } y_{kuv_i} > 0\}$
9: Let's define a probability function $Prob$ which assigns for each v_i probability equal to $\frac{y_{kuv_i}}{\sum_{v_j \in R_u} y_{kuv_j}}$
10: Choose randomly $v_i \in R_u$, where probability of selecting an element v_i equals $Prob(v_i)$
11: Add vertex v_i to P_k
12: $u = v_i$
13: **end while**
14: **end for**
15: Add constraints $\{P_k : k \in S\}$ to the $problem$
16: Solve the $problem$
17: Retrieve $\{x_k : k \in S\}$ and U from the $problem$ solution
18: Update FAS with values $\{(x_k, P_k) : k \in S\}$
19: **return** $(FAS, \sum_{k\in S} u(x_k))$

Algorithm 4 LPBestPathNum(G, c, F_{st})

1: Create a $problem$ with goal as (13)
2: Add constraints {(2),(4),(5),(6),(7),(14),(15)} to the $problem$
3: Solve the $problem$
4: Retrieve $\{y_{(k,u,v)\in S\times V\times V}\}$ and U from the $problem$ solution
5: **for** $k \in S$ **do**
6: Using the BFS based procedure choose the max path P_k with value y_{P_k} (max min y_{kuv} on the path)
7: **end for**
8: Add constraints $\{P_k : k \in S\}$ to the $problem$
9: Solve the $problem$
10: Retrieve $\{(x_k, P_k) : k \in S\}$ and U from the $problem$ solution
11: Update FAS with values $\{(x_k, P_k) : k \in S\}$
12: **return** $(FAS, \sum_{k\in S} u(x_k))$

a constraint (random selection, since there may be more than one best path (line 13)). The constraint forces i-flow to use only P_i edges in the next iterations.
In the last $|S|+1$ iteration the LP relaxation is solved assuming single fixed paths for all the $k \in S$ flows (line 17).

F. Mixed integer linear program NUM

The algorithm *MilpNum* is based on *MILP formulation* of the problem. It is a single phase algorithm (line 3).

Algorithm 5 IterativeLPBestPathNum(G, c, F_{st})

1: Create a $problem$ with goal as (13)
2: Add constraints {(2),(4),(5),(6),(7),(14),(15)} to the $problem$
3: Let's initiate a set of users with fixed single path $S' = \emptyset$
4: $stop = false$
5: **while** $(stop == false)$ **do**
6: **if** $(S \setminus S'$ is not $\emptyset)$ **then**
7: Solve the $problem$
8: Retrieve $\{y_{(k,u,v)\in S\times V\times V}\}$ and U from the $problem$ solution
9: **for** $i \in S \setminus S'$ **do**
10: Using the BFS based procedure choose the max path P_i with value y_{P_i} (max min y_{iuv} on the path)
11: **end for**
12: Let's define a set M as following $M = \{i \in S \setminus S'\text{: } y_{P_i} \cdot x_i == max\{y_{P_j} \cdot x_j\}_{j\in S\setminus S'}\}$
13: Choose randomly $i \in M$, where probability of selecting an element i equals $\frac{1}{|M|}$
14: Add constraint P_i to the $problem$
15: $S' = S' \cup \{i\}$
16: **else**
17: Solve the $problem$
18: $stop = true$
19: **end if**
20: **end while**
21: Retrieve $\{(x_k, P_k) : k \in S\}$ and U from the $problem$ solution
22: Update FAS with values $\{(x_k, P_k) : k \in S\}$
23: **return** $(FAS, \sum_{k\in S} u(x_k))$

Algorithm 6 MilpNum(G, c, F_{st})

1: Create a $problem$ with goal as (13)
2: Add constraints {(2),(4),(5),(6),(7),(8),(9),(10),(14)} to the $problem$
3: Solve the $problem$
4: Retrieve $\{(x_k, P_k) : k \in S\}$ and U from the $problem$ solution
5: Update FAS with values $\{(x_k, P_k) : k \in S\}$
6: **return** $(FAS, \sum_{k\in S} u(x_k))$

VI. Computational results

A. Experiment Setup

The algorithms were programmed in Python 2.7 with the use of CPLEX Python library 12.8.0. The algorithm *MilpNum* uses CPLEX executable. All the simulations were run on a PC with 2-core 2.7GHz and 8GB RAM.

The experiments were conducted on five types of networks. The first network *Net7* with 7 nodes and 10 links [9]. The second network *Net22* with 22 nodes and 40 links [2]. In the second network nodes 1-8 and T (which is a network exit node) were source-destination nodes, whereas nodes 9-

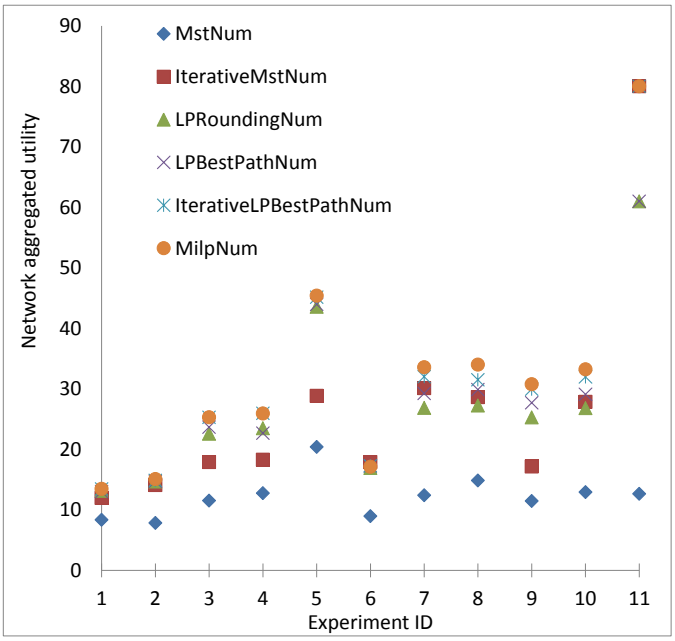


Fig. 3. Network aggregate utility (experiments $S1 - S11$)

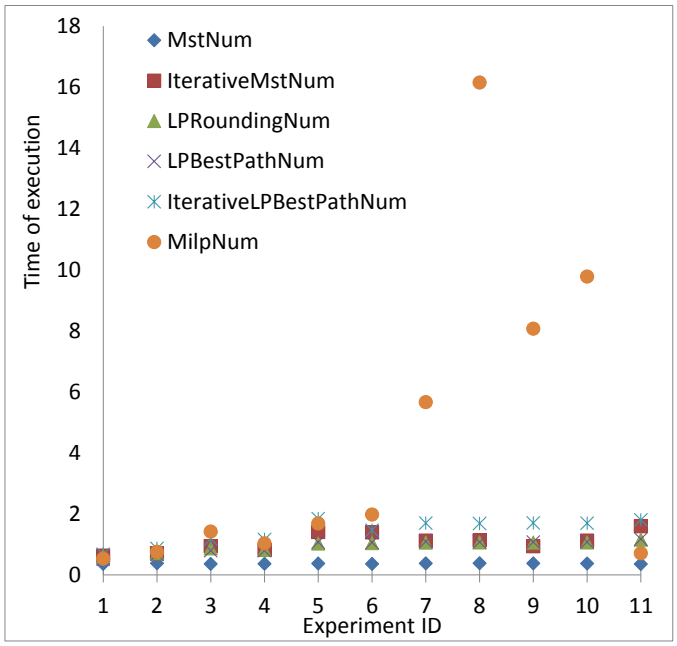


Fig. 4. Time of execution $\sqrt[4]{T}$ [sec.] (experiments $S1 - S11$)

21 were transit nodes. The third, grid network *Net64* with 8x8 nodes. The fourth network *Net42* with 40 disjoint paths spanned between node 1 and 42. Finally, the fifth, grid network *Net400* with 20x20 nodes. In all the networks above each link had a capacity equal to 1.

1) Verification of algorithm performance for different network structure and size: The experiment parameters were as follows: S1 - Net7, 10 pairs as given in [9]; S2 - Net22, 10 users randomly selected [1]; S3/4 - Net22, 20 users randomly selected [1], S5 - Net22, 50 users randomly selected [1]; S6 - Net64, 20 users: 10 x (1,64) and 10 x (8,57); S7/8 - Net64, 20 users randomly selected; S9/10 - Net64, 20 users randomly selected [2]; S11 - Net42, 20 users (1,42).

The algorithm *IterativeMstNum* was run with $\Delta_\varepsilon = 0.1$.

In each experiment, the algorithms: *LPRoundingNum*, *LPBestPathNum* and *IterativeLPBestPathNum* were run 20 times and the average values of objective U_{avg} and execution time T_{avg} are presented.

2) Verification of algorithm performance for different number of users: Additional 10 experiments (on Net64) were conducted to demonstrate an increase in utility and time upon an increase in the number of users: S12 - 10, S13 - 20, ... , S21 - 100 users [2].

3) Verification of algorithm performance for a large network: Finally, additional 3 experiments were conducted to capture utility and time for the Net400: S22 - 20, S23 - 50, S24 - 100 users [2].

[1] 50% of the pairs constitute internal traffic (between nodes 1-8), and 50% of the pairs constitute external traffic (between nodes 1-8 and T)

[2] Traffic sources and targets were located on the grid edge, and were randomly selected. Each smaller user set was a subset of a larger set.

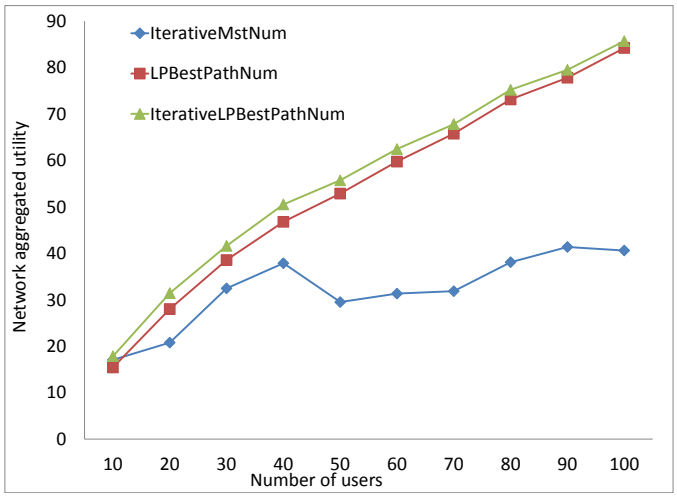


Fig. 5. Network aggregate utility in function of number of users (experiments $S12 - S21$)

B. Performance of IterativeMstNum heuristics

1) Different network structure and size - experiments $S1 - S11$: ***IterativeMstNum* vs *MstNum*.** The results from $S1 - S11$ demonstrate a large improvement in the objective value of the algorithm *IterativeMstNum* in comparison to *MstNum* (Fig. 3) - efficiency (utility) was even several times higher as high.

***IterativeMstNum* vs *LPRoundingNum*.** The utility values of *IterativeMstNum* observed in $S1 - 5, 9$ were not as high as the average ones achieved by *LPRoundingNum* (utility of *LPRoundingNum* was higher by $4\% - 51\%$ (Fig. 3)). On the other hand, in $S6 - 8, S10 - 11$ the results were in favor of *IterativeMstNum* (utility of *IterativeMstNum* was higher by

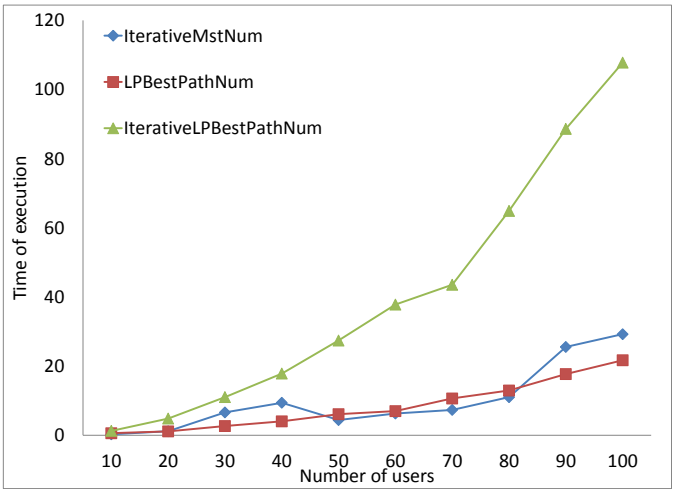


Fig. 6. Time of execution [sec.] in function of number of users (experiments $S12 - S21$)

TABLE I
UTILITY AND TIME FOR A LARGE NETWORK (EXPERIMENTS $S22 - S24$)

Experiment ID	Number of users	IterativeMstNum		LPBestPathNum		IterativeLPBestPathNum	
		U	$T[s]$	U_{avg}	$T_{avg}[s]$	U_{avg}	$T_{avg}[s]$
S22	20	33,36	10,96	31,54	27,09	36,18	139,78
S23	50	38,85	30,34	63,59	341,51	77,46	6152,44
S24	100	36,49	208,56	109,40	5129,93	-	-

$4\% - 31\%$ (Fig. 3)).

When it comes to the observed execution time, the algorithms *IterativeMstNum* and *LPRoundingNum* exposed similar performance, slightly in favour of *LPRoundingNum* (Fig. 4).

***IterativeMstNum* vs *MilpNum*.** The objective values of *IterativeMstNum* observed in the experiments $S6, S11$ were very close to the ones achieved by *MilpNum*. In the experiment $S6$ the value of *MilpNum*, lower than the value of *IterativeMstNum*, can be explained by an approximation error (Fig. 3). For a small number of users ($S1, S2$) the observed distance from the optimum (*MilpNum*) was not more than 12% , for a larger number of users ($S3 - S5, S7 - 10$) it ranged between 11% and 80%.

2) Different number of users - experiments $S12 - S21$: ***IterativeMstNum* vs *LPBestPathNum*.** In general, the observed utility of *IterativeMstNum* increased when the number of users grew, however the function was not monotonic. The utility was comparable with the one achieved by *LPBestPathNum* when the number of users was not high, up to 40 users for *Net64* (Fig. 5). Then, the utility difference was not higher than 35%. When the number of users was 10, for *Net64*, the utility of *IterativeMstNum* was higher by 11%. This number of users was correlated with the number of user-disjoint paths.

The time of execution of *IterativeMstNum* was comparable with the one achieved by *LPBestPathNum* regardless of the number of users (Fig. 6).

3) Large network - experiments $S22 - S24$: ***IterativeMstNum* vs *LPBestPathNum*.** The utility of *IterativeMstNum* was comparable with the one achieved by *LPBestPathNum* when the number of users was not high (TABLE I). When number of users was 20, for *Net400*, the utility of *IterativeMstNum* was higher by 5.8%. The same remark on user-disjoint paths applies, as above.

The time of execution of *IterativeMstNum* was much smaller than the one achieved by *LPBestPathNum* (an order of magnitude when the number of users was high - TABLE I).

C. Performance of LPBestPathNum and IterativeLPBestPathNum heuristics

1) Different network structure and size - experiments $S1 - S11$: ***LPBestPathNum* vs *LPRoundingNum*.** The average objective values of *LPBestPathNum* observed in the most of the experiments $S1 - S11$ were superior to *LPRoundingNum* by up to 9.7% (Fig. 3), except for $S1, S4$ - 1% and 3.6% opposite difference, respectively. Both heuristics demonstrated very similar execution time (Fig. 4).

***LPBestPathNum* vs *IterativeLPBestPathNum*.** In $S1 - S11$ the average observed objective values of *IterativeLPBestPathNum* were higher than values of *LPBestPathNum* by up to 14.5%, except for the parallel edge network (*Net42*) by 31% (Fig. 3). However, *IterativeLPBestPathNum* exposed higher execution time due to additional $|S| - 1$ iterations (Fig. 4).

***IterativeLPBestPathNum* vs *MilpNum*.** In $S1 - S6, S11$ the optimal objective values (*MilpNum*) were higher than *IterativeLPBestPathNum* solution by not more than 1.5%. In $S7, S9 - 10$ - by not more than 4.8%. In $S8$ - by 7.9%.

2) Different number of users - experiments $S12 - S21$: ***IterativeLPBestPathNum* vs *LPBestPathNum*.** In $S12 - S21$ the average objective values of *IterativeLPBestPathNum* and *LPBestPathNum* grew when the number of users grew. The utilities of *IterativeLPBestPathNum* were higher than those of *LPBestPathNum* (Fig. 5). The observed difference became smaller when the number of user grew. As observed, from 16% for 10 users to 2% for 100 users.

However, *IterativeLPBestPathNum* required much longer computation times than *LPBestPathNum*, with an increasing number of users (Fig. 6).

3) Large network - experiments $S22 - S24$: ***IterativeLPBestPathNum* vs *LPBestPathNum*.** The utility difference between *IterativeLPBestPathNum* and *LPBestPathNum* was more visible for large networks (20x20 nodes), as observed 14.7% (20 users) - 21.8% (50 users) in favor of the former (TABLE I). However, the computation time of *IterativeLPBestPathNum* for a large networks and large number of users may not be practically attainable (TABLE I).

D. Performance of MilpNum solver

The experiments $S7 - 10$ showed that *MilpNum* may expose a very long computation time to achieve an optimal solution,

in the range of hours (Fig. 4). A possible, way around for *MilpNum* would be to configure the automatic termination when reaching a defined solution gap. As an example, in the experiment $S9$ *MilpNum* gave the value 26,22 (gap 16,95%) in 12s and 29 (5,75%) in 90s. IterativeLPBestPathNum gave 29,928 in 8,4s. In $S10$ *MilpNum* gave 30 (11,11%) in 12s and 31,66 (5%) in 268s. IterativeLPBestPathNum gave 31,962 in 8,2s.

VII. Conclusions

Since our problem is NP-hard, search for optimum (*MilpNum*) may not be practically attainable even for medium-sized grid networks (8x8 nodes) and a small number of users (~ 20).

The experiments show that the algorithm *LPBestPathNum* demonstrates a reasonably good trade-off between computing time and precision of the utility value.

As observed, our LP-based heuristics *IterativeLPBestPathNum* demonstrates up to $\sim 15\%$ higher utility, on networks with up to 64 nodes and 50 users, but it is slower, in comparison to *LPBestPathNum*. The difference in utility may be a few percent bigger for large grid networks (20x20 nodes) and a high number of users (~ 50), and even more for a special type of network - parallel edges. The bigger difference is caused by the edge allocation conflict. The *IterativeLPBestPathNum* reduces the probability of such conflicts by iterative path fixing (a single path fixed for each iteration).

The utility values of *IterativeLPBestPathNum*, on network with up to 64 nodes and 50 users, are lower than *MilpNum* by only $\sim 2\%$ on average.

The successful *IterativeLPBestPathNum* computations were conducted for grid networks with up to 20x20 nodes and 50 users. For comparison, the successful and not terminated *MilpNum* computations were conducted for grid networks with up to 8x8 nodes and 20 users.

As observed, our MST-based heuristics *IterativeMstNum* ensures reasonably good utility in certain cases. When the number of users is not high (the number is correlated with the number of user-disjoint paths), the *IterativeMstNum* utility is lower than *MilpNum* by not more than $\sim 10\%$. With such an assumption it has efficiency comparable to *LPBestPathNum*. When the number of users grows, the gap between *IterativeMstNum* and *LPBestPathNum* increases. The *IterativeMstNum* has utility which is equal to optimal for a special type of network - parallel edge network, since it avoids allocating the same edge to competing streams if alternative ones are available. In such networks, it may be much more efficient than *LPBestPathNum* (e.g. $S11 - 30\%$).

On average, for the observed cases, the *IterativeMstNum* utility value is lower than *MilpNum* by $\sim 25\%$.

Additionally, *IterativeMstNum* has computing time comparable to *LPBestPathNum* for small and medium-sized networks. For large networks and a high number of users, it may be an order of magnitude faster than *LPBestPathNum*. Finally, it does not use LP solver but a simple MST-based iterative procedure.

References

[1] A. Amokrane, R. Langar, R. Boutaba, and G. Pujolle. Flow-based management for energy efficient campus networks. *IEEE Transactions on Network and Service Management*, 12(4):565–579, Dec 2015. 10.1109/TNSM.2015.2501398.

[2] A.P. Bianzino, L. Chiaraviglio, M. Mellia, and J.L. Rougier. Grida: Green distributed algorithm for energy-efficient ip backbone networks. *Computer Networks*, 56(14):3219–3232, September 2012. https://doi.org/10.1016/j.comnet.2012.06.011.

[3] M. Chiang, S. H. Low, A. R. Calderbank, and J. C. Doyle. Layering as optimization decomposition: A mathematical theory of network architectures. *Proceedings of the IEEE*, 95(1):255–312, Jan 2007. 10.1109/JPROC.2006.887322.

[4] M. Drwal. Approximation algorithms for utility-maximizing network design problem. In Henry Selvaraj et al., editor, *Progress in Systems Engineering: Proceedings of the Twenty-Third International Conference on Systems Engineering*, volume 366 of *Advances in Intelligent Systems and Computing*. Springer International Publishing, Cham, 2015. https://doi.org/10.1007/978-3-319-08422-0-60.

[5] S. Even, A. Itai, and A. Shamir. On the complexity of time table and multi-commodity flow problems. In IEEE, editor, *16th Annual Symposium on Foundations of Computer Science 13-15 October 1975*, October 1975. 10.1109/SFCS.1975.21.

[6] L. K. Fleischer. Approximating fractional multicommodity flow independent of the number of commodities. In *40th Annual Symposium on Foundations of Computer Science (Cat. No.99CB37039)*, pages 24–31, Oct 1999. 10.1109/SFFCS.1999.814573.

[7] N. Garg and J. Konemann. Faster and simpler algorithms for multicommodity flow and other fractional packing problems. In *Proceedings 39th Annual Symposium on Foundations of Computer Science (Cat. No.98CB36280)*, pages 300–309, Nov 1998. 10.1109/SFCS.1998.743463.

[8] X. Huang, T. Yuan, and M. Ma. Utility-optimized flow-level bandwidth allocation in hybrid sdns. *IEEE Access*, 6:20279–20290, 2018. 10.1109/ACCESS.2018.2820682.

[9] P. Jaskóła, P. Arabas, and A. Karbowski. Simultaneous routing and flow rate optimization in energy-aware computer networks. *International Journal of Applied Mathematics and Computer Science*, 26(1):231–243, 2016. 10.1515/amcs-2016-0016.

[10] F.P. Kelly, A.K. Maulloo, and D.K.H. Tan. Rate control for communication networks: shadow prices, proportional fairness and stability. *Journal of the Operational Research Society*, 49(3):237–252, March 1998. https://doi.org/10.1057/palgrave.jors.2600523.

[11] J.W. Lee, M. Chiang, and R. A. Calderbank. Jointly optimal congestion and contention control based on network utility maximization. *IEEE Communications Letters*, 10(3):216–218, March 2006. 10.1109/LCOMM.2006.1603389.

[12] T. Leighton and S. Rao. Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms. *Journal of the ACM*, 46(6):787–832, November 1999.

[13] M. Macit, V. C. Gungor, and G. Tuna. Comparison of qos-aware single-path vs. multi-path routing protocols for image transmission in wireless multimedia sensor networks. *Ad Hoc Networks*, 19:132 – 141, 2014. https://doi.org/10.1016/j.adhoc.2014.02.008.

[14] J. Mo and J. Walrand. Fair end-to-end window-based congestion control. *IEEE/ACM Transactions on Networking*, 8(5):556–567, Oct 2000. 10.1109/90.879343.

[15] M. E. Pfetsch and Ch. Liebchen. Multicommodity flows and column generation. In *Lecture Notes*. Zuse Institute Berlin, 2006.

[16] P. R. Satav and P. M. Jawandhiya. Review on single-path multi-path routing protocol in manet: A study. In *2016 International Conference on Recent Advances and Innovations in Engineering (ICRAIE)*, pages 1–7, Dec 2016. 10.1109/ICRAIE.2016.7939514.

[17] J. Wang, L. Li, S.H. Low, and J.C. Doyle. Cross-layer optimization in tcp/ip networks. *IEEE/ACM Transactions on Networking*, 13(3):582 – 595, June 2005. 10.1109/TNET.2005.850219.

[18] M. Wang, C. W. Tan, W. Xu, and A. Tang. Cost of not splitting in routing: Characterization and estimation. *IEEE/ACM Transactions on Networking*, 19(6):1849–1859, Dec 2011. 10.1109/TNET.2011.2150761.

[19] W. Wang, M. Palaniswami, and S. H. Low. Application-oriented flow control: Fundamentals, algorithms and fairness. *IEEE/ACM Transactions on Networking*, 14(6):1282–1291, Dec 2006. 10.1109/TNET.2006.886318.

Advances in Computer Science & Systems

CSS 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—12th Workshop on Computer Aspects of Numerical Algorithms
- C&SS'19—6th International Conference on Cryptography and Security Systems
- LTA'19—4th International Workshop on Language Technologies and Applications
- MMAP'19—12th International Symposium on Multimedia Applications and Processing
- WAPL'19 7th Workshop on Advances in Programming Languages
- WSC'19—10th Workshop on Scalable Computing

12th Workshop on Computer Aspects of Numerical Algorithms

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

TOPICS

- Parallel numerical algorithms
- Novel data formats for dense and sparse matrices
- Libraries for numerical computations
- Numerical algorithms testing and benchmarking
- Analysis of rounding errors of numerical algorithms
- Languages, tools and environments for programming numerical algorithms
- Numerical algorithms on coprocesors (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,** Universita' 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

Communication Papers of the Federated Conference on Computer Science and Information Systems pp. 41–45
DOI: 10.15439/2019F164
ISSN 2300-5963 ACSIS, Vol. 20

Symbolic description of the polynomial roots and their numerical implementation – better than in Mathematica software?

Edyta Hetmaniok
Damian Słota
Mariusz Pleszczyński
Roman Wituła
and Michał Różański
Silesian University of Technology
Institute of Mathematics
Email: edyta.hetmaniok@polsl.pl
mariusz.pleszczynski@polsl.pl
roman.witula@polsl.pl
michal.rozanski@polsl.pl

Marcin Szczygieł
Silesian University of Technology
Institute of Mechatronics
Email: marcin.szczygiel@polsl.pl

***Abstract*—This paper is a continuation of the discussion undertaken in paper [31]. We present in the current paper the corrected, and also given in a slightly changed form, Vandermonde formulae for the roots of some quintic polynomials considered in J.P. Tignol's monograph [26]. The proofs of selected trigonometric identities from [31] are given and some new identities have been generated by the occasion, which also can be used for testing our Langrange algorithm for the case of cubic polynomials. Moreover, we present here the decomposition of polynomials belonging to some two-parameter family of polynomials related to the Chebyshev polynomials of the first kind.**

Keywords – Lagrange algorithm, Chebyshev polynomials

I. Introduction

In paper [31] we have presented an algorithm (the so called Lagrange algorithm) for determining the complex roots of real polynomials which appeared to be more effective than the symbolic calculations proposed in Mathematica software. Many cubic polynomial, considering the polynomials used for testing our Lagrange algorithm, led us to some attractive trigonometric relations (see the appendix which is an important supplement of paper [31]). In the current paper we propose to repeat our previous considerations for the case of quintic polynomials and for the polynomials of higher degrees as well.

Let us also recall that the complex roots of quintic polynomials and the polynomials of higher degree, described with the aid of elliptic functions and hypergeometric functions, are well known (see [2], [12]). Certainly, the computational usefulness of such solutions is not good, it still remains in the sphere of future discoveries. At present, we have to limit the calculations, especially the symbolic calculations, just for the selected families of polynomials, for which the formulae describing roots are well known and are in the "numerically accepted" form.

II. Vandermonde formulae

Next of the investigated quintic polynomials

$$x^5 - x^4 - 4x^3 + 3x^2 + 3x - 1 = \prod_{k=1}^{5} \left(x + 2\cos\frac{2k\pi}{11} \right)$$

delivered to us many unexpected emotions. So in monograph [26] the following identity, based on the Lagrange resolvent method of solving the algebraic equations, is suggested

$$x = \frac{1}{5}\left(1 + \Delta' + \Delta'' + \Delta''' + \Delta^{IV}\right), \tag{1}$$

where x is one of the roots of this polynomial. However the description of expressions Δ', Δ'', Δ''' and Δ^{IV}, given in [26], "seems to be" incorrect. It should be, and this is already our correction of these relations: $\Delta' = \sqrt[5]{d_1}$, $\Delta'' = \overline{\Delta'}$, $\Delta''' = \sqrt[5]{d_3}$, $\Delta^{IV} = \overline{\Delta'''}$, more precisely, Δ' and Δ''' are any values of the respective complex roots of degree five taken from numbers (respectively):

$$d_1 = \frac{11}{4}\left(89 + 25\sqrt{5} - 5i\sqrt{5 + 2\sqrt{5}} + 45i\sqrt{5 - 2\sqrt{5}}\right),$$

$$d_3 = \frac{11}{4}\left(89 - 25\sqrt{5} - 45i\sqrt{5 + 2\sqrt{5}} - 5i\sqrt{5 - 2\sqrt{5}}\right).$$

Identity (1) takes then the following trigonometric form[1] (the calculations were executed by hand with the intensive support

[1]We used in calculations the following identity

$$d\sqrt{a + b\sqrt{a}} + \sqrt{a - b\sqrt{a}} = \sqrt{(d^2+1)a + (b(d^2-1) + 2cd)\sqrt{a}},$$

where $a = b^2 + c^2$, $b, c \in \mathbb{C}$ (the complex cases need some additional verification).

©2019, PTI

of Mathematica software):

$$-2\cos\frac{2k\pi}{11}=\frac{1}{5}\left(1+2\sqrt{11}\cos\left(\frac{1}{5}\left(t_i+s_i\arctan\frac{\Im d_1}{\Re d_1}\right)\right)\right.$$
$$\left.+\sqrt{11}\cos\left(\frac{1}{5}\left(t_j-r_j\arctan\frac{\Im d_3}{\Re d_3}\right)\right)\right)$$

and hence, after simple algebraic transformations, we get

$$-5\cos\frac{2k\pi}{11}-\frac{1}{2}=\sqrt{11}\left(\cos\left(\frac{1}{5}\left(t_i+s_i\arctan\frac{p}{q}\right)\right)\right.$$
$$\left.+\cos\left(\frac{1}{5}\left(t_j+r_j\arctan\frac{pq}{109}\right)\right)\right),\quad(2)$$

where $p=5\sqrt[4]{5}$, $q=\sqrt{877\sqrt{5}+1958}$, s_i, r_i are the signs and $t_n:=2(n-1)\pi$ for every $n=1,2,\ldots,5$. By using the Mathematica software we have found the following sequences of the values of indices k,i,j and the signs s_i, r_j for which both the above identities hold true:

k	i	j	s_i	r_j	k	i	j	s_i	r_j
1	3	3	-1	-1	3	1	4	$+1$	-1
1	3	4	-1	$+1$	3	1	4	-1	-1
1	4	3	$+1$	-1	4	2	2	$+1$	-1
1	4	4	$+1$	$+1$	4	2	5	$+1$	$+1$
2	3	2	$+1$	$+1$	4	5	2	-1	-1
2	3	5	$+1$	-1	4	5	5	-1	$+1$
2	4	2	-1	$+1$	5	2	1	-1	-1
2	4	5	-1	-1	5	2	1	-1	$+1$
3	1	3	$+1$	$+1$	5	5	1	$+1$	-1
3	1	3	-1	$+1$	5	5	1	$+1$	$+1$

III. Quintic Polynomials

In this section we discuss the monic irreducible quintic polynomials $f\in\mathbb{Q}[x]$. If the equation $f(x)=0$ is solvable by radicals, the quintic polynomial $f(x)$ is said to be solvable. First let us consider the quintic polynomial in the reduced form

$$x^5+a_2x^2+a_1x+a_0=0,$$

called the principal quintic form, and

$$x^5+b_1x+b_0=0,$$

called the Bring-Jerrard quintic form (see [1]).

It is known that if $f(x)=x^5+ax^i+b\in\mathbb{Q}^*[x]$, $i=1,2$, where $\mathbb{Q}^*:=\mathbb{Q}\setminus\{0\}$, and disc $f(x)>0$[2], then $f(x)$ has exactly one real root [24].

B.K. Spearman and K.S. Williams proved in [23] the following result.

Theorem 1. *Let $a,b\in\mathbb{Q}^*$ be such that the quintic trinomial $f(x)=x^5+ax+b$ is irreducible. Then $f(x)$ is solvable if and only if there exist the rational numbers $\varepsilon\in\{-1,1\}$, $c\geq 0$ and $\alpha\neq 0$ such that*

$$a=\frac{5\alpha^4(3-4\varepsilon c)}{c^2+1},\quad b=\frac{-4\alpha^5(11\varepsilon+2c)}{c^2+1}.$$

Moreover, if we set

$$D=c^2+1,$$

[2]The discriminant of f.

$$\begin{cases}v_1=\sqrt{D}+\sqrt{D-\varepsilon\sqrt{D}}, & v_2=-\sqrt{D}-\sqrt{D+\varepsilon\sqrt{D}},\\ v_3=-\sqrt{D}+\sqrt{D+\varepsilon\sqrt{D}}, & v_4=\sqrt{D}-\sqrt{D-\varepsilon\sqrt{D}},\end{cases}$$

$$u_1=\sqrt[5]{\frac{v_1^2v_3}{D^2}},\ u_2=\sqrt[5]{\frac{v_3^2v_4}{D^2}},\ u_3=\sqrt[5]{\frac{v_2^2v_1}{D^2}},\ u_4=\sqrt[5]{\frac{v_4^2v_2}{D^2}},$$

and $w=\exp(2\pi i/5)$, then the roots of x^5+ax+b are equal to x_0, x_1, x_2, x_3, x_4, where

$$\frac{x_i}{\alpha}=\sum_{k=1}^{4}w^{ki}u_k$$
$$=w^iu_1+\frac{w^{2i}u_1^2\sqrt{D}\sqrt[5]{\varepsilon}}{v_1}+\frac{w^{3i}u_1^3D}{v_1v_3}-\frac{w^{4i}u_1^4\sqrt{D^3}}{v_1^2v_3}\sqrt[5]{\frac{\varepsilon}{D}},$$

for every $i=0,1,2,3,4$.

For illustrating the application of the formulae from the above theorem let us find the complex roots of polynomial

$$x^5+11x+44,$$

which we get for parameters $\varepsilon=1$, $\alpha=-1$, $c=2/11$. By using the above formulae, supported by Mathematica software, we derive the following formulae

$$x_0=\lambda(\xi_1-\xi_2-\xi_3-\xi_4)\approx-1.87775,$$
$$x_1=-\lambda\left(e^{\frac{3\pi i}{5}}\xi_1+e^{\frac{2\pi i}{5}}\xi_2+e^{\frac{4\pi i}{5}}\xi_3-e^{\frac{\pi i}{5}}\xi_4\right)$$
$$\approx 1.80012-1.44746i,$$
$$x_2=-\lambda\left(e^{\frac{\pi i}{5}}\xi_1+e^{\frac{4\pi i}{5}}\xi_2-e^{\frac{3\pi i}{5}}\xi_3+e^{\frac{2\pi i}{5}}\xi_4\right)$$
$$\approx-0.861241-1.9105i,$$
$$x_3=\bar{x_2},\quad x_4=\bar{x_1},$$

where $\lambda=-\sqrt[5]{11/5^4}$ and

$$\xi_k:=\left((-1)^{k+1}(50\sqrt{5}-75)+\sqrt{18125-6169\sqrt{5}}\right)^{1/5},$$

$$\xi_l:=\left(50\sqrt{5}+75+(-1)^l\sqrt{18125+6169\sqrt{5}}\right)^{1/5},$$

for $k=1,2$ and $l=3,4$.

J.A. Johnstone and B.K. Spearman in paper [11] proved the following result.

Theorem 2. *Let $f(x)=x^5+x-a$, where $a\in\mathbb{Z}$. Then $f(x)$ is not solvable by radicals unless $a=r^5+r$ for some integer r, or $a=\pm1,\pm6$. The last case was discussed and proved earlier by S. Rabinowitz [21].*

B.K. Spearman and K.S. Williams in paper [24] proved also the completely unexpected result, given below.

Theorem 3. *There exist only five essentially different, irreducible, solvable, quintic trinomials x^5+ax^2+b, $a,b\in\mathbb{Q}^*$, namely: x^5+5x^2+3, x^5+5x^2-15, x^5+25x^2+300, x^5+100x^2+1000 and x^5+250x^2+625.*

IV. Appendix

In paper [31] we have presented the following formulae

$$\left(z-2\sin\frac{4\pi}{13}\right)\left(z-2\sin\frac{10\pi}{13}\right)\left(z-2\sin\frac{12\pi}{13}\right) = z^3-\sqrt{\frac{13+3\sqrt{13}}{2}}z^2+\sqrt{13}z-\sqrt{\frac{13-3\sqrt{13}}{2}} \tag{3}$$

$$\left(z-2\sin\frac{2\pi}{13}\right)\left(z-2\sin\frac{6\pi}{13}\right)\left(z+2\sin\frac{8\pi}{13}\right) = z^3-\sqrt{\frac{13-3\sqrt{13}}{2}}z^2-\sqrt{13}z+\sqrt{\frac{13+3\sqrt{13}}{2}} \tag{4}$$

with the suggestion that they can be derived on the basis of formulae given below

$$z^3+\frac{1-\sqrt{13}}{2}z^2-z+\frac{\sqrt{13}+3}{2} = \left(z-2\cos\frac{2\pi}{13}\right)\left(z-2\cos\frac{6\pi}{13}\right)\left(z-2\cos\frac{8\pi}{13}\right), \tag{5}$$

$$z^3+\frac{1+\sqrt{13}}{2}z^2-z-\frac{\sqrt{13}+3}{2} = \left(z-2\cos\frac{4\pi}{13}\right)\left(z-2\cos\frac{10\pi}{13}\right)\left(z-2\cos\frac{12\pi}{13}\right). \tag{6}$$

To explain it better we decided to "reveal" here the details of the proof. Thus, by proving decomposition (3) and (4) we used the following relations

$$\begin{aligned}8\sin\frac{4\pi}{13}\sin\frac{10\pi}{13}\sin\frac{12\pi}{13} &= i\left(e^{i\frac{4\pi}{13}}-e^{-i\frac{4\pi}{13}}\right)\\ &\times\left(e^{i\frac{10\pi}{13}}-e^{-i\frac{10\pi}{13}}\right)\left(e^{i\frac{12\pi}{13}}-e^{-i\frac{12\pi}{13}}\right)\\ &= 2\sin\frac{2\pi}{13}+2\sin\frac{6\pi}{13}-2\sin\frac{8\pi}{13},\end{aligned}$$

$$\begin{aligned}&\left(2\sin\frac{4\pi}{13}+2\sin\frac{10\pi}{13}+2\sin\frac{12\pi}{13}\right)^2\\ &= 6+2\left(\cos\frac{2\pi}{13}+\cos\frac{6\pi}{13}+\cos\frac{8\pi}{13}\right)\\ &-4\left(\cos\frac{4\pi}{13}+\cos\frac{10\pi}{13}+\cos\frac{12\pi}{13}\right)\overset{(5)}{\underset{(6)}{=}}\frac{13+3\sqrt{13}}{2},\end{aligned}$$

$$\begin{aligned}&4\left(\sin\frac{4\pi}{13}\sin\frac{10\pi}{13}+\sin\frac{4\pi}{13}\sin\frac{12\pi}{13}+\sin\frac{10\pi}{13}\sin\frac{12\pi}{13}\right)=\\ &= 2\left(\cos\frac{2\pi}{13}+\cos\frac{6\pi}{13}+\cos\frac{8\pi}{13}\right)\\ &-2\left(\cos\frac{4\pi}{13}+\cos\frac{10\pi}{13}+\cos\frac{12\pi}{13}\right)\overset{(5)}{\underset{(6)}{=}}\sqrt{13},\end{aligned}$$

$$\begin{aligned}&8\sin\frac{2\pi}{13}\sin\frac{6\pi}{13}\sin\frac{8\pi}{13}\\ &= i\left(e^{i\frac{2\pi}{13}}-e^{-i\frac{2\pi}{13}}\right)\left(e^{i\frac{6\pi}{13}}-e^{-i\frac{6\pi}{13}}\right)\left(e^{i\frac{8\pi}{13}}-e^{-i\frac{8\pi}{13}}\right)\\ &= 2\sin\frac{4\pi}{13}+2\sin\frac{10\pi}{13}+2\sin\frac{12\pi}{13},\end{aligned}$$

$$\begin{aligned}&\left(2\sin\frac{2\pi}{13}+2\sin\frac{6\pi}{13}-2\sin\frac{8\pi}{13}\right)^2\\ &= 6+2\left(\cos\frac{4\pi}{13}+\cos\frac{10\pi}{13}+\cos\frac{12\pi}{13}\right)\\ &-4\left(\cos\frac{2\pi}{13}+\cos\frac{6\pi}{13}-\cos\frac{8\pi}{13}\right)\overset{(5)}{\underset{(6)}{=}}\frac{13-3\sqrt{13}}{2},\end{aligned}$$

$$\begin{aligned}&4\left(\sin\frac{2\pi}{13}\sin\frac{6\pi}{13}-\sin\frac{2\pi}{13}\sin\frac{8\pi}{13}-\sin\frac{6\pi}{13}\sin\frac{8\pi}{13}\right)\\ &= 2\left(\cos\frac{4\pi}{13}+\cos\frac{10\pi}{13}+\cos\frac{12\pi}{13}\right)\\ &-2\left(\cos\frac{2\pi}{13}+\cos\frac{6\pi}{13}+\cos\frac{8\pi}{13}\right)\overset{(5)}{\underset{(6)}{=}}-\sqrt{13}.\end{aligned}$$

In order to complete the collection of formulae (3)–(6) we give additionally the decompositions connected with the values of tangent and cotangent functions of the respective ternary sets of arguments $\left\{\frac{2\pi}{13},\frac{6\pi}{13},\frac{8\pi}{13}\right\}$ and $\left\{\frac{4\pi}{13},\frac{10\pi}{13},\frac{12\pi}{13}\right\}$. They are of the following form

$$\begin{aligned}&\left(x-\tan\frac{4\pi}{13}\right)\left(x-\tan\frac{10\pi}{13}\right)\left(x-\tan\frac{12\pi}{13}\right)\\ &= x^3-\sqrt{65-18\sqrt{13}}x^2+(13-4\sqrt{13})x-\sqrt{65-18\sqrt{13}},\end{aligned}$$

since from formulae (3)–(6) we easily obtain the equalities

$$\begin{aligned}\tan\frac{4\pi}{13}\tan\frac{10\pi}{13}\tan\frac{12\pi}{13} &= \frac{2\sin\frac{4\pi}{13}2\sin\frac{10\pi}{13}2\sin\frac{12\pi}{13}}{2\cos\frac{4\pi}{13}2\cos\frac{10\pi}{13}2\cos\frac{12\pi}{13}}\\ &= \frac{\sqrt{\frac{13-3\sqrt{13}}{2}}}{\frac{\sqrt{13}+3}{2}} = \frac{\sqrt{13}-3}{4}\sqrt{26-6\sqrt{13}} = \sqrt{65-18\sqrt{13}},\end{aligned}$$

$$\tan\frac{4\pi}{13}+\tan\frac{10\pi}{13}+\tan\frac{12\pi}{13} = \tan\frac{4\pi}{13}\tan\frac{10\pi}{13}\tan\frac{12\pi}{13},$$ [3]

[3]The following identity holds

$$\tan x+\tan y+\tan(k\pi-x-y) = \tan x\tan y\tan(k\pi-x-y)$$

for the respective values of $x,y\in\mathbb{R}$ and $k\in\mathbb{Z}$, which in turn implies the identity

$$\cot x\cot y+\cot x\cot(k\pi-x-y)+\cot y\cot(k\pi-x-y) = 1.$$

Moreover, if $x_1+x_2+x_3 = 2\pi$, $x_1,x_2,x_3\in\mathbb{R}\setminus\pi\mathbb{Z}$, then the following identities hold

$$\begin{aligned}&-2\left(\cot x_1+\cot x_2+\cot x_3\right)\sin x_1\sin x_2\sin x_3\\ &\quad = \sin^2 x_1+\sin^2 x_2+\sin^2 x_3 = (\sin x_1+\sin x_2+\sin x_3)^2\\ &\quad -2\sin x_1\sin x_2-2\sin x_1\sin x_3-2\sin x_2\sin x_3.\end{aligned}$$

Proof. We have

$$\cot x_1+\cot x_2 = -\frac{\sin^2 x_3}{\sin x_1\sin x_2\sin x_3},$$

$$\cot x_1+\cot x_3 = -\frac{\sin^2 x_2}{\sin x_1\sin x_2\sin x_3},$$

$$\cot x_2+\cot x_3 = -\frac{\sin^2 x_1}{\sin x_1\sin x_2\sin x_3},$$

from which, by summing respectively by sides, we obtain the expected equality. □

the last formula implies the following one

$$\cot\frac{4\pi}{13}+\cot\frac{10\pi}{13}+\cot\frac{12\pi}{13}=\frac{\frac{1}{4}\frac{13+3\sqrt{13}}{2}-\frac{\sqrt{13}}{2}}{-\frac{1}{4}\sqrt{\frac{13-3\sqrt{13}}{2}}}$$
$$=-\frac{\sqrt{2}}{4}\sqrt{13+3\sqrt{13}}(\sqrt{13}-1)=-\frac{\sqrt{2}}{2}\sqrt{26+4\sqrt{13}},$$

$$\tan\frac{4\pi}{13}\tan\frac{10\pi}{13}+\tan\frac{4\pi}{13}\tan\frac{12\pi}{13}+\tan\frac{10\pi}{13}\tan\frac{12\pi}{13}$$
$$=\tan\frac{4\pi}{13}\tan\frac{10\pi}{13}\tan\frac{12\pi}{13}\left(\cot\frac{4\pi}{13}+\cot\frac{10\pi}{13}+\cot\frac{12\pi}{13}\right)$$
$$=\sqrt{65-18\sqrt{13}}\left(-\frac{\sqrt{2}}{2}\sqrt{26+4\sqrt{13}}\right)$$
$$=-\sqrt{\left(65-18\sqrt{13}\right)\left(13+2\sqrt{13}\right)}=13-4\sqrt{13}.$$

The next formulae, derived by us on the way of algebraic tests of the Lagrange algorithm, are as follows (we present them without the proofs with respect to the length of this paper):

$$\left(x-\cot\frac{4\pi}{13}\right)\left(x-\cot\frac{10\pi}{13}\right)\left(x\cot\frac{12\pi}{13}\right)$$
$$=x^3+\frac{\sqrt{2}}{2}\sqrt{26+4\sqrt{13}}x^2+x-\sqrt{\frac{1}{13}(65+18\sqrt{13})},$$

$$\left(x-\tan\frac{2\pi}{13}\right)\left(x-\tan\frac{6\pi}{13}\right)\left(x+\tan\frac{8\pi}{13}\right)=x^3-$$
$$\sqrt{65-18\sqrt{13}}x^2+(13+4\sqrt{13})x-\sqrt{65-18\sqrt{13}},$$

$$\left(x-\cot\frac{2\pi}{13}\right)\left(x-\cot\frac{6\pi}{13}\right)\left(x+\cot\frac{8\pi}{13}\right)$$
$$=x^3-\sqrt{13-2\sqrt{13}}x^2+x-\sqrt{\frac{1}{13}(65-18\sqrt{13})},$$

$$\left(x-2\cos\frac{2\pi}{21}\right)\left(x-2\cos\frac{8\pi}{21}\right)\left(x-2\cos\frac{32\pi}{21}\right)$$
$$=\left(x-2\cos\frac{2\pi}{21}\right)\left(x-2\cos\frac{8\pi}{21}\right)\left(x+2\cos\frac{11\pi}{21}\right)$$
$$=x^3-\frac{1+\sqrt{21}}{2}x^2-\frac{1-\sqrt{21}}{2}x-\frac{5-\sqrt{21}}{2},$$

$$\left(x-2\cos\frac{4\pi}{21}\right)\left(x-2\cos\frac{16\pi}{21}\right)\left(x-2\cos\frac{64\pi}{21}\right)$$
$$=\left(x-2\cos\frac{4\pi}{21}\right)\left(x+2\cos\frac{5\pi}{21}\right)\left(x+2\cos\frac{\pi}{21}\right)$$
$$=x^3-\frac{1-\sqrt{21}}{2}x^2-\frac{1+\sqrt{21}}{2}x-\frac{5+\sqrt{21}}{2},$$

since $\cos\frac{16\pi}{21}=-\cos\frac{5\pi}{21}$, $\cos\frac{32\pi}{21}=\cos\frac{10\pi}{21}=-\cos\frac{11\pi}{21}$ and $\cos\frac{64\pi}{21}=-\cos\frac{\pi}{21}=\cos\frac{20\pi}{21}$. Finally the last formula is of the form

$$\prod_{k=1}^{6}\left(x-2\cos\frac{2^k\pi}{21}\right)=$$
$$=\left(x^3-\frac{1+\sqrt{21}}{2}x^2-\frac{1-\sqrt{21}}{2}x-\frac{5-\sqrt{21}}{2}\right)$$
$$\times\left(x^3-\frac{1-\sqrt{21}}{2}x^2-\frac{1+\sqrt{21}}{2}x-\frac{5+\sqrt{21}}{2}\right)$$
$$=x^6-x^5-6x^4+6x^3+8x^2-8x+1.$$

Furthermore we have (see [29]):

$$\left(x-2\sin\frac{2\pi}{21}\right)\left(x-2\sin\frac{8\pi}{21}\right)\left(x-2\sin\frac{32\pi}{21}\right)$$
$$=\left(x-2\sin\frac{2\pi}{21}\right)\left(x-2\sin\frac{8\pi}{21}\right)\left(x+2\sin\frac{10\pi}{21}\right)$$
$$=x^3+\frac{\sqrt{3}-\sqrt{7}}{2}x^2-\frac{3+\sqrt{21}}{2}x+\frac{\sqrt{3}+\sqrt{7}}{2},$$

$$\left(x-2\sin\frac{4\pi}{21}\right)\left(x-2\sin\frac{16\pi}{21}\right)\left(x-2\sin\frac{64\pi}{21}\right)$$
$$=\left(x-2\sin\frac{4\pi}{21}\right)\left(x-2\sin\frac{5\pi}{21}\right)\left(x+2\sin\frac{\pi}{21}\right)$$
$$=x^3-\frac{\sqrt{3}+\sqrt{7}}{2}x^2+\frac{\sqrt{21}-3}{2}x+\frac{\sqrt{7}-\sqrt{3}}{2},$$

since $\sin\frac{16\pi}{21}=\sin\frac{5\pi}{21}$, $\sin\frac{32\pi}{21}=-\sin\frac{10\pi}{21}=-\sin\frac{11\pi}{21}$ and $\sin\frac{64\pi}{21}=-\sin\frac{\pi}{21}=-\sin\frac{20\pi}{21}$, which implies

$$\prod_{k=1}^{6}\left(x-2\sin\frac{2^k\pi}{21}\right)$$
$$=x^6-\sqrt{7}x^5-2x^4+4\sqrt{7}x^3-8x^2+1.$$

V. Special families of polynomials

There exist many special families of polynomials with the well known sets of roots. However there is no any common platform connecting these families and this is, to be honest, a great challenge of our undertaking. Considering the special families of polynomials, discovered by us, we want to emphasize these ones, the origins of which can be found in Cardano's formulae (see for example [28]). So if $\Omega_n(x)$ denote the so called n-th Chebyshev polynomial defined by the relation

$$\Omega_n(x):=2T_n\left(\frac{x}{2}\right),$$

where $T_n(\cos\theta)=\cos n\theta$ is the n-th Chebyshev polynomial of the first kind, then the following decompositions hold

$$(\sqrt{p})^n\Omega_n\left(\frac{x}{\sqrt{p}}\right)-q=\prod_{k=0}^{n-1}\left(x-a\zeta^{2k}-b\zeta^{-2k}\right),$$

$$(-\sqrt{-p})^n\Omega_n\left(\frac{ix}{\sqrt{p}}\right)+q=\prod_{k=0}^{n-1}\left(x-a\zeta^{2k+1}+b\zeta^{-2k-1}\right),$$

where $\zeta = \exp(i\pi/n)$, a and b are any values from among the n complex values of the following roots (respectively):

$$\sqrt[n]{\frac{1}{2}\left(q+\sqrt{q^2-4p^n}\right)}, \quad \sqrt[n]{\frac{1}{2}\left(q-\sqrt{q^2-4p^n}\right)}.$$

These formulae lead to many special cases, they generate many trigonometric identities. However the disadvantage of these identities is their dependence only on two complex parameters p and q, no matter what is the value of the discussed polynomial degree. We also observe some calculating difficulties – however just on the same kind as in case of the classical Cardano's formulae.

VI. Conclusions

Our attempt of making the algorithms, serving for symbolic determination of the complex roots of some real polynomials, more effective brought us many practical advantages. We have obtained the algorithms behaving better, at least in the cases emphasized by us, in comparison with, for example, the procedures used in Mathematica software. By the way we have generated many original identities and we have discovered new relations. We have also learned to avoid verious traps appearing in the course of using the algebra of complex numbers. Considering the future plans we intend still to seek the families of polynomials with the known description of roots. Also the form of realizing our mega-algorithm, which should end our future investigations, still remains questionable.

References

[1] V.S. Adamchik and D.J. Jeffrey, "Polynomial Transformations of Tschirnhaus, Bring and Jerrard", *ACM SIGSAM Bulletin*, vol. 37, no 3, 2003, pp. 90–94.

[2] G. Belardinelli, *Fonctions hypergéométriques de plusieurs variables et résolution analytique des équations algébriques générales*, Gauthier-Villars, Paris 1960.

[3] B.C. Berndt, B.K. Spearman and K.S. Williams, "Commentary on an unpublished lecture by G.N. Watson on solving the quintic", *Math. Intelligencer*, vol. 24, 2002, pp. 15–33.

[4] D.S. Dummit, "Solving solvable quintics", *Math. Comp.*, vol. 57, 1991, pp. 387–401.

[5] M. Elia and P. Filipponi, "Equations of the Bring-Jerrard form, the golden section, and square fibonacci numbers", *The Fibonacci Quarterly*, vol. 36 no. 3, 1998, pp. 282–286.

[6] H. Funato and A. Kawamura, "Control of variable active-passive reactance (VAPAR) and negative inductance", *PESC '94 Record., 25th Annual IEEE*, pp. 189–196 vol.1, ISBN 0-7803-1859-5, DOI 10.1109/PESC.1994.349731.

[7] J.C. Glashan, "Notes on the quintic", *Amer. J. Math.*, vol. 7, no. 2, 1885, pp. 178–179.

[8] L. Guishu, D. Huaying, W. Tao and C. Xiang, "Generalized Kuroda's identity and its applications in nonuniform transmission lines", *Proceedings. 2003 6th ISAPE*, pp. 839–842, ISBN 0-7803-7831-8, DOI 10.1109/ISAPE.2003.1276817.

[9] T.R. Hagedorn, "General formulas for solving solvable sextic equations", *J. Algebra*, vol. 233, no. 2, 2000, pp. 704–751.

[10] E. Hetmaniok, P. Lorenc, M. Pleszczyński and R. Wituła, "Iterated integrals of polynomials", *Appl. Math. Comput.*, vol. 249, 2014, pp. 389–398.

[11] J.A. Johnstone and B.K. Spearman, "On a sequence of nonsolvable quintic polynomials", *J. Integer Seq.*, vol. 12, no. 2, 2009, Art. ID 09.2.8.

[12] F. Klein, *Vorlesungen Über das Ikosaeder und die Auflösung der Gleichungen vom Fünften Grade*, Leipzig 1884 (we used the Russian translation: Nauka Press, Moscow 1989).

[13] S. Landau, "$\sqrt{2}+\sqrt{3}$: four different views", *The Math. Intelligencer*, vol. 20, no. 4, 1998, pp. 55–60.

[14] S. Landau and G.L. Miller, "Solvability by radicals is in polynomial time", *J. Comp. System Sciences*, vol. 30, 1985, pp. 179–208.

[15] C.A. Leal-Sevillano, J.R. Montejo-Garai, J.A. Ruiz-Cruz and J.M. Rebollar, "Wideband Equivalent Circuit for Multi-Aperture Multi-Resonant Waveguide Irises", *IEEE TMTT*, vol. 64, iss. 3, pp.724–732, ISSN 0018-9480, DOI 10.1109/TMTT.2016.2520462.

[16] E. Mc Clintock, "Further researches in the theory of quintic equations", *Amer. J. Math.*, vol. 20, no. 2, 1898, pp. 157–192.

[17] A. Mostowski and M. Stark, *Elements of the higher algebra*, PWN, Warszawa 1972 (in Polish).

[18] L.J. Okunev, *Ring of polynomials and field of rational functions*, in Mathematics, its subject, methods and significance, Part II, Academy of Science Press, Moscow 1951 (in Russian).

[19] F.W.I. Olver, D.W. Lozier, R.F. Boisvert and C.W. Clark, *NIST Handbook of Mathematical Functions*, Cambridge Univ. Press, 2010.

[20] A.P. Prudnikov, A.J. Bryczkov and O.I. Mariczev, *Integrals and series*, Elementary Functions, Nauka, Moscow 1981 (in Russian).

[21] S. Rabinowitz, "The factorization of $x^5 \pm x + n$", *Math. Magazine*, vol. 61, 1988, pp. 191–193.

[22] B. Resmana, R.P. Astuti and A. Kurniawan, *Wiener filter-based channel predictor performance improvement using polinomial extrapolation*, 3rd ICICI-BME, Bandung 2013, pp.184–189, ISBN 978-1-4799-1649-8, DOI 10.1109/ICICI-BME.2013.6698489.

[23] B.K. Spearman and K.S. Williams, "Characterization of solvable quintics $x^5 + ax + b$", *Amer. Math. Mon.*, vol. 101, no 10, 1994, pp. 986–992.

[24] B.K. Spearman and K.S. Williams, "On solvable quintics $x^5 + ax + b$ and $x^5 + ax^2 + b$", *Rocky Mountain J. Math.*, vol. 26, 1996, pp. 753–772.

[25] B.K. Spearman and K.S. Williams, "The factorization of $x^5 \pm x^a + n$", *The Fibonacci Quarterly*, vol. 36, 1998, pp. 158–170.

[26] J.-P. Tignol, *Galois' Theory of Algebraic Equations*, World Scientific, New Jersey 2001.

[27] R. Wituła, E. Hetmaniok, D. Słota and N. Gawrońska, "Sums of the rational powers of roots of the polynomials", *International Journal of Pure and Applied Mathematics*, vol. 85, no. 1, 2013, pp. 179–191.

[28] R. Wituła and D. Słota, "Cardano's formula, square roots, Chebyshev polynomials and radicals", *J. Math. Anal. Appl.*, vol. 363, 2010, pp. 639–647.

[29] R. Wituła and D. Słota, "Fixed and periodic points of polynomials generated by minimal polynomials of $2\cos(2\pi/n)$", *Int. J. Bifur. Chaos*, vol. 19, no. 9, 2009, pp. 1–12.

[30] R. Wituła and D. Słota, "On modified Chebyshev polynomials", *J. Math. Anal. Appl.*, vol. 324, 2006, pp. 321–343. Corrigendum to "On modified Chebyshev polynomials", *J. Math. Anal. Appl.*, vol. 336, 2007, p. 750.

[31] A. Wróbel, E. Hetmaniok, M. Pleszczyński and R. Wituła, "On an improvement of the numerical application for Cardano's formula in Mathematica software", *Proc. Inter. Symp. Young Scien. Techn., Eng. Math. SYSTEM 2015*, Catania, Italy, 2015, pp. 27–29, online: http://ceur-ws.org/Vol-1543/p10.pdf.

12th International Symposium on Multimedia Applications and Processing

ORGANIZED 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 (MMAP 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 MMAP 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

MMAP 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 MMAP 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. MMAP 2016 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 Extractio
- 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
- **Korzhik, 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

Communication Papers of the Federated Conference on Computer Science and Information Systems pp. 49–54
DOI: 10.15439/2019F182
ISSN 2300-5963 ACSIS, Vol. 20

The Use of Modern Multimedia Trends for Popularization of Mechatronics by Presenting Its Present and Future Application

Erik Kučera, Oto Haffner, Roman Leskovský, Jakub Matišák and Peter Drahoš
Faculty of Electrical Engineering and Information Technology
Slovak University of Technology in Bratislava
Bratislava, Slovakia
Email: erik.kucera@stuba.sk

***Abstract*—Mechatronics belongs to the modern fields of the science and technology. Modern mechatronics combines mechanics, electronics, automation and informatics. However, young people often understand mechatronics as a combination of mechanics and electronics. This erroneous paradigm brings many problems in education and industry. Actual trend in the industry is Industry 4.0 - also called as fourth industrial revolution. It is a collective name for current automation, exchanging of data and manufacturing technologies. Today's industry (digital factories) is full of information and communication technologies such as Internet of Things, virtual / mixed reality, cloud computing, etc. Unfortunately, only small amount of high school pupils would like to study mechatronics because young generation is influenced by the outdated perception of this field and does not perceive some of its new aspects. They do not see modern technologies in this field. The authors of the paper decided to present mechatronics as a modern branch in an original way - by a fanfilm inspired by Avengers: Infinity War.**

I. Introduction

IN THE past, the industry was affected by technological change and innovation. These paradigms are called industrial revolutions. These revolutions were caused by mechanization (1st industrial revolution), use of electrical energy (2nd industrial revolution) and electronics and automation (3rd industrial revolution). All these industrial revolutions did not influence only the production itself, but also the labour market and the educational system as well. As a result of these changes some professions and jobs disappeared. Currently, due to the development of digitalization and robotics, we are facing the next industrial revolution, known as the Industry 4.0. It is expected that some professions will be replaced. The emerging technologies have huge effect on the education of people. Only qualified and highly educated employees will be able to control these technologies. The industry should collaborate with universities. The main vision the following of Industry 4.0 is the emergence of "smart factories", that will be connected to the production facilities Cyber-physical systems called CPS. Using of the Internet of Things, the Internet of Services and the Internet of People will make connection: machine-machine, human-machine or human-human, and at the same time an enormous amount of data will be obtained [1]. For this reason it will be necessary to analyse large data

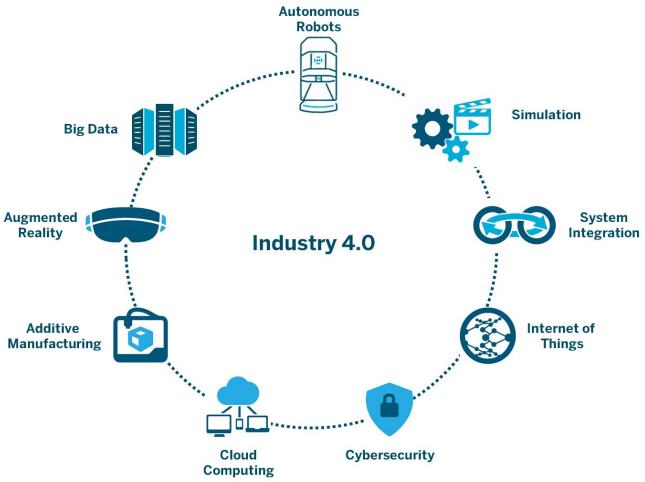


Fig. 1. Modern technologies for Industry 4.0 [4]

(Big Data) to be able to predict possible failures and adapt in real time to the changed conditions [2].

Now, let's analyze the Industry 4.0 impact on mechatronics. This should include the state-of-the-art in research, education and development of control methods, control structures, information a communications technologies. Their applications in different types of industrial processes with the focus on mechatronic systems and the new trends declared in Industry 4.0 should be considered, as well [3].

Mechatronics merges methodology and basic principles of four close scientific disciplines: automation, information and communication technologies (ICT) and mechanics. Thereby it represents an excellent example of a multidisciplinary engineering and points out how an integration of disciplines leads to new degrees of freedom in the corresponding new forms of education [3] [5].

Multidisciplinary education in mechatronics for Industry 4.0 enables an engineer to perform initial design work in mechanical, electrical, automation, control engineering and information technologies and to communicate effectively with

©2019, PTI

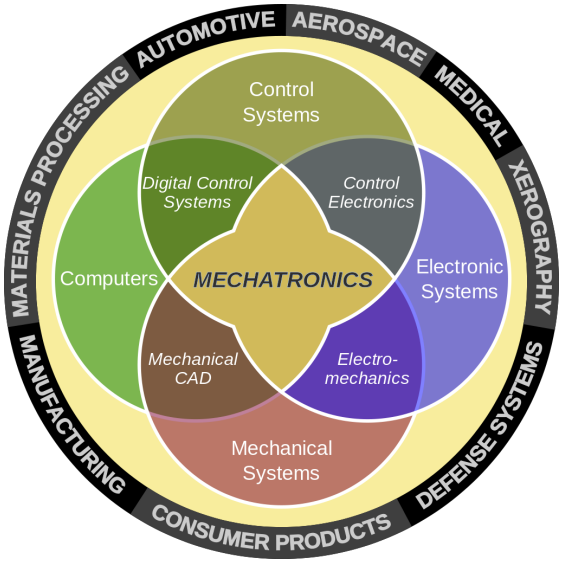


Fig. 2. Mechatronics diagram

specialized design engineers in different disciplines (mechanical, electrical, ICT, power, etc.). Universities need to shape their engineering curricula to better prepare students for professional practice [3].

Nowadays, modern technologies such as cloud, artificial intelligence, Internet of Things, virtual and augmented reality, etc., are used in mechatronics. According to the definition, mechatronics synergistically combines informatics, automation, mechanics, and electronics into a single integrated whole (Fig. 2). To get a mechatronic device, the 4 fields and the individual components of the device have to be effectively interconnected; the binder are smart technologies. Still, smart technologies cannot be considered as a fifth extra field of mechatronics but rather a "soul" of the mechatronic product without which its individual components would not work as a whole.

Unfortunately, only small amount of high school pupils would like to study mechatronics because young generation is influenced by the outdated perception of this field and does not perceive some of its new aspects. They do not see modern or information and communication technologies in this field. As it was stated, this is not true at all. The authors of the paper decided to present mechatronics as a modern branch in an original way.

II. Popularization of Mechatronics Using Multimedia

The authors decided to inspire with the very successful promotional videos of the universities that were inspired by the popular film theme (Wolf from VUT - Wolf from Wall Street [6] or Join the FEI STU - Star Wars [7].

Nowadays, modern technologies or multimedia are used in education or popularization of science [8] [9] [10] [11] [12].

During discussions and polls for high school students, it was found that high school students do not know what to imagine under the field of mechatronics. They have often lived with very old understanding of mechatronics as a combination of mechanics and electronics or electrical engineering. This is not mechatronics but electromechanics. It is therefore necessary to draw the young generation's attention to this modern engineering field in a popular way with the expected viral effect. It is more and more challenging to take interest of the young generation (influenced by high-budget production of films and series from Hollywood) in audiovisual work, and especially in education or science and technology. The more emphasis has to be placed on the production quality. Today's young people are used to the high quality video and the aforementioned high-budget production and will not be satisfied with the average audiovisual quality of the final product. Therefore, the use of a well-known theme can support an interest in viewing such a work, which, however, must be processed at an adequate level.

The authors decided to make a fanfilm inspired by Avengers: Infinity War [13]. The aim of the film is to introduce mechatronics to the young generation in an attractive way as a modern 21st century discipline that combines informatics, automation, mechanics and electronic systems. In the fanfilm, the parallel between the aspects of the universe (= Infinity Stones from Avengers: Infinity War) and the 4 parts of mechatronics (= Mechatronics Stones) are used. The fifth stone is a source of tension and surprise moment.

Many industrial and service equipment as well as consumer products are based on the combination of mechanical and electronic components. However, machines, devices, equipment and products are becoming mechatronic only when they have some degree of "intelligence" as a result of the modern development of electronics, information-communication and control technologies (automation).

As it was stated, to get a mechatronic device, the 4 fields and the individual components of the device have to be effectively interconnected; the binder are smart technologies. So , the fifth stone represents smart technologies.

Our fanfilm was named as Avengers: Five Mechatronics Stones (in Slovak: Avengeri: Päť kameňov mechatroniky). It has about 10 minutes.

III. Preproduction of Fanfilm

The comparison between the Infinity Stones and the fields of mechatronics area came as a very interesting idea for a promotional video. We had filmed more promotional and education videos at the Institute of Automotive Mechatronics at the Faculty of Electrical Engineering and Information Technology, Slovak University of Technology in Bratislava. But these videos were simpler and shorter. Thus, the story was written, and the screenplay was gradually developed and the filming started. Since the beginning, we have realized that this is our most comprehensive project in the last 10 years.

Restrictions and challenges:

- *The use of personnel capital at the Institute* - We wanted to make every element of the fanfilm in-house. All participants are from the mentioned Institute or other institutes of our faculty.

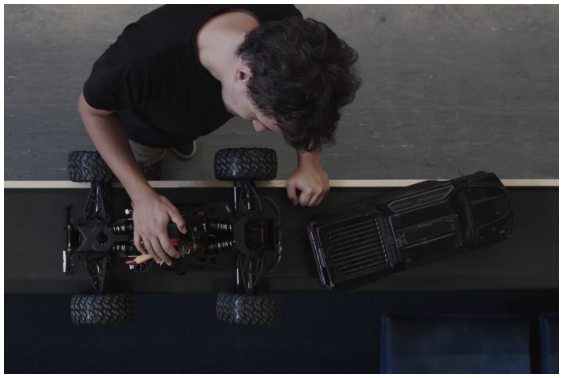


Fig. 3. RC model of car

- *Trying to cast girls* - It was a huge problem to find willing girl protagonists.
- *The use of faculty spaces* - They are often not representative and appropriate for popularization and propagation.
- *Time restrictions* - It was filmed mainly when the semester lessons were not in progress. Even so, it was difficult to match the time schedules of the actors. One of the actors has to be added additionally using chroma-keying techniques.
- *Won't it be too long?* - Should we keep the complete Marvel Cinematic Universe film structure (2 credits + mid-credits scene + after-credits scene)?
- *"Zero" budget* - The whole project had almost zero budget, colleagues helped us with the passion for work, we used only existing devices, laboratories, video editing software, etc. We only bought a few things inspired by the Marvel Cinematic Universe. Our PhD student is called Erich Stark (resembles Tony Stark in original films), so of course we used this name creatively. Over time, we realized that another of the PhD students was a bit like Captain America, so we also included this in the video.

We shot the film during seven or eight shooting days. Most of the time it took to prepare a thoughtful screenplay to fit everything together, and especially the very demanding postproduction, as the film contains a large number of special visual effects. Preproduction started practically a few days after the premiere of Avengers: Infinity War. So the whole production lasted about 7-8 months during regular scientific and pedagogical work.

IV. Fanfilm Analysis

The described fanfilm can be seen here (English subtitles can be turned on): https://youtu.be/RtROkZqU6ns or http://avengers.mechatronika.cool/.

A. Introduction in space

In the introduction, by microphone voice individual Mechatronics Stones are explained and also what is mechatronics.

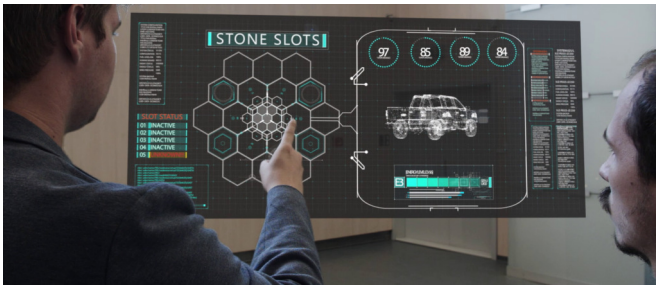


Fig. 4. Slots of Mechatronics Stones

Fig. 5. Chroma-keying

B. Why do we need Mechatronics Stones?

The main idea of the film is to use stones to make the mechatronic device work. As a device, a mechatronic vehicle model was selected (Fig. 3). Five Mechatronics Stones must be inserted into the virtual slots (Fig. 4). The slots have been filmed using chroma-keying techniques (Fig. 5).

C. Electronics Stone

The first stone to be introduced is Electronics Stone (Fig. 7). This section shows the work with electronics (Fig. 6). The result of working with electronics is a programmable LED cube.

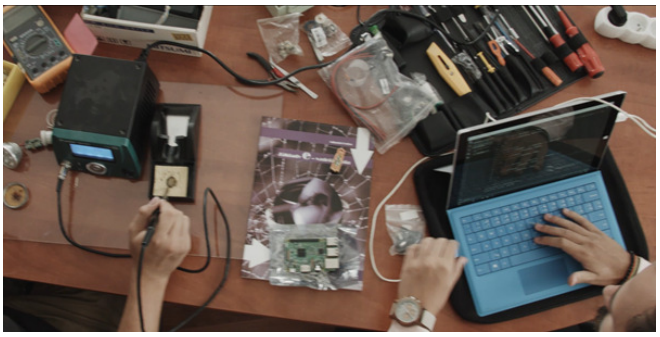


Fig. 6. The work with electronics

Fig. 7. Electronics Stone

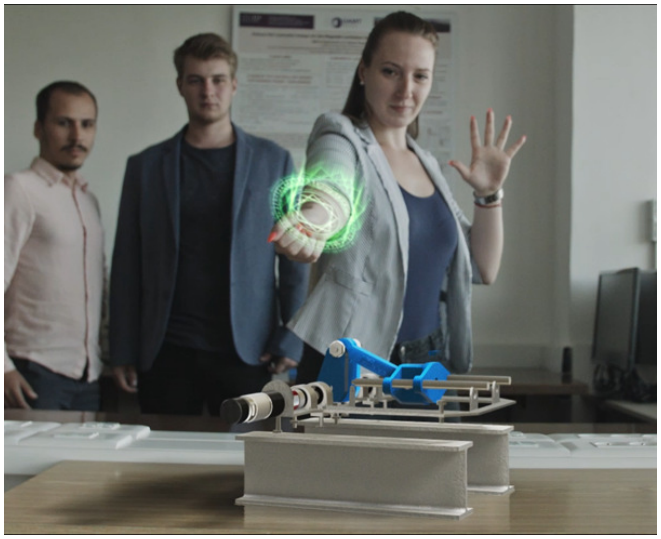


Fig. 8. The use of Mechanics Stone

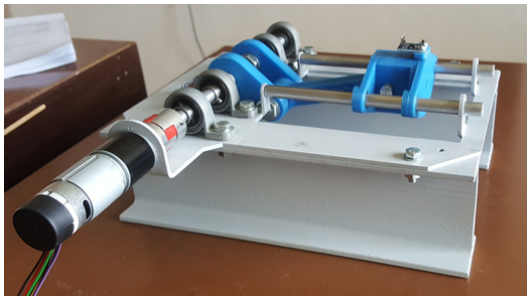


Fig. 9. Real mechanical device

Fig. 10. Microsoft HoloLens and augmented / mixed reality

Fig. 11. Iconic shot of Erich Stark inspired by Iron Man

D. Mechanics Stone

The second stone is the Mechanics Stone. It is inspired by the Time Stone from original films. Using the Mechanics Stone, a device is constructed (Fig. 8). At our institute, the displayed device is really available (Fig. 9).

E. Informatics Stone

The third stone is the Informatics Stone. Our PhD student - Erich Stark (Fig. 11) - is introduced. Informatics is presented by augmented / mixed reality and by holographic device Microsoft HoloLens (10).

F. Automation Stone

The last part of mechatronics is automation. Automation is presented by a model of magnetic levitation (Fig. 12). The Automation Stone is hidden in a ball controlled by magnets (Fig. 13).

G. Smart Technologies Stone

The protagonists found that with four Mechatronics Stones the device did not work (Fig. 14). Smart technologies are needed to link fields of mechatronics. Smart technologies Stone is located in glasses (Fig. 15) of a new hero. Glasses are a symbol of intelligence. The stones are joined and the device works (Fig. 16).

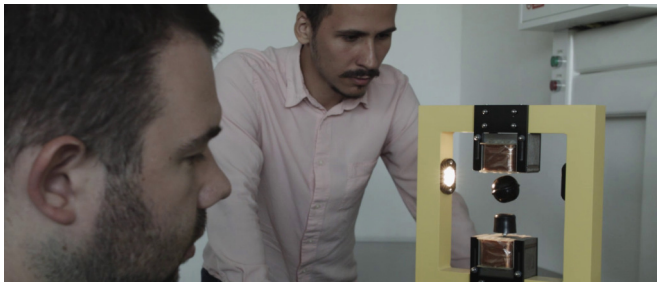


Fig. 12. Magnetic levitation model

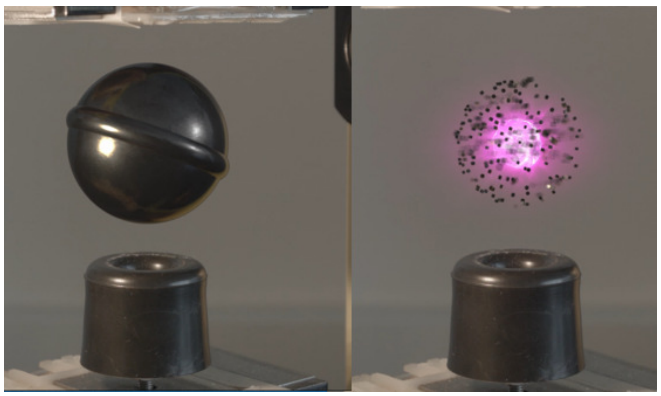


Fig. 13. Automation Stone

H. The end and mid-credits scene

The new hero turns to be an antagonist. He use the stones and snaps his fingers. Then our protagonists are turned into the dust (Fig. 17).

V. Critical Response

The described fanfilm achieved critical success (Fig. 18). It has about 14,000 views on 14/05/2019 on YouTube. Many Slovak media published articles about fanfilm: Nový Čas (online and print) [14], Refresher [15], Start it up [16], etc. Also, about fanfilm was written abroad [17].

Fig. 14. The device does not work

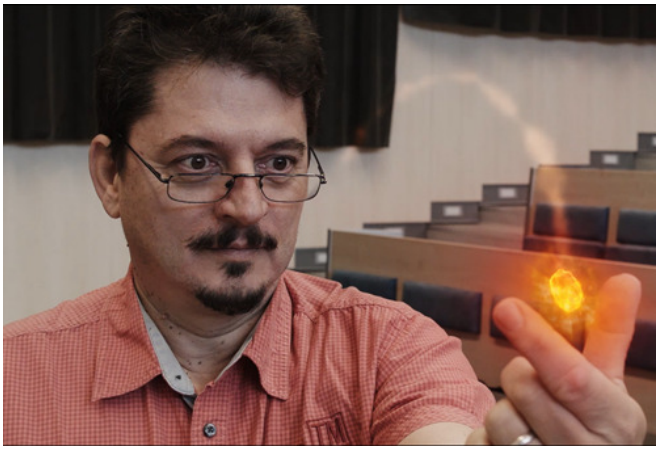


Fig. 15. Smart technologies Stone is located in glasses

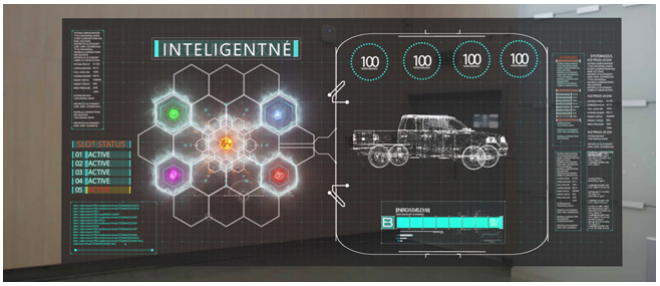


Fig. 16. All Mechatronics Stones

Fig. 17. Mid-credits scene

Fig. 18. Likes and dislikes on YouTube (14/05/2019)

VI. Conclusion

Multidisciplinary education in mechatronics for Industry 4.0 enables an engineer to perform initial design work in mechanical, electrical, automation, control engineering and information technologies and to communicate effectively with specialized design engineers in different disciplines (mechanical, electrical, ICT, power, etc.). Nowadays, modern technologies such as cloud, artificial intelligence, Internet of Things, virtual and augmented reality, etc., are used in mechatronics. Unfortunately, only small amount of high school pupils would like to study mechatronics because young generation is influenced by the outdated perception of this field and does not perceive some of its new aspects. The authors of the paper decided to present mechatronics as a modern branch in an original way - by the fanfilm inspired by Avengers: Infinity War. The described popularization fanfilm achieved critical success and appeared in mass media.

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. 1324 – VZRI4 (Virtual and Mixed Reality for 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] A. Benešová and J. Tupa, "Requirements for education and qualification of people in industry 4.0," *Procedia Manufacturing*, vol. 11, pp. 2195 – 2202, 2017. doi: https://doi.org/10.1016/j.promfg.2017.07.366 27th International Conference on Flexible Automation and Intelligent Manufacturing, FAIM2017, 27-30 June 2017, Modena, Italy. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2351978917305747

[3] M. Huba and Š.. Kozák, "From e-learning to industry 4.0," in *2016 International Conference on Emerging eLearning Technologies and Applications (ICETA)*, Nov 2016. doi: 10.1109/ICETA.2016.7802083 pp. 103–108.

[4] T. Melanson. (2018) What industry 4.0 means for manufacturers. [Online]. Available: https://aethon.com/mobile-robots-and-industry4-0/

[5] S. Kozak, E. Ruzicky, J. Stefanovic, and F. Schindler, "Research and education for industry 4.0: Present development," in *2018 Cybernetics & Informatics (K&I)*, Feb 2018, pp. 1–7.

[6] M. Balík and et al. (2014) Wolf of but (in czech). [Online]. Available: https://youtu.be/cCNpNpni6rY

[7] J. Seewald and et al. (2017) Join the fei stu (in slovak). [Online]. Available: https://youtu.be/xrF6L3Uc1Ik

[8] 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.

[9] 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

[10] J. Majernik, M. Madar, and J. Mojzisova, "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

[11] K. Szklanny, Ł. Homoncik, M. Wichrowski, and A. Wieczorkowska, "Creating an interactive and storytelling educational physics app for mobile devices," in *2017 Federated Conference on Computer Science and Information Systems (FedCSIS)*, Sep. 2017. doi: 10.15439/2017F95 pp. 1269–1273.

[12] R. R. Gajewski, "Pitfalls of e-education: From multimedia to digital dementia?" in *2016 Federated Conference on Computer Science and Information Systems (FedCSIS)*, Sep. 2016, pp. 913–920.

[13] Marvel. (2018) Avengers: Infinity war. [Online]. Available: https://www.marvel.com/movies/avengers-infinity-war

[14] A. Galajdová. (2019) Stu students want to convince that their field is not as boring as many think: You will see the video, you will be amazed! (in slovak). [Online]. Available: https://www.cas.sk/clanok/811257/studenti-z-stu-chcu-presvedcit

[15] J. Paulík. (2019) Slovak students shooted a film inspired by the avengers (in slovak). [Online]. Available: https://filmkult.refresher.sk/13207-Slovenski-vysokoskolaci-natocili-film-inspirovany-Avengers

[16] V. Andorová. (2019) Slovak students shooted a short film inspired by the avengers for almost zero costs (in slovak). [Online]. Available: https://bit.ly/2vZ48qm

[17] D. Hanuš. (2019) Avengers: Five mechatronics stones or marvel's competition in bratislava (in czech). [Online]. Available: https://bit.ly/30gx37f

Advances in Network Systems and Applications

MODERN 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—3rd Workshop on Internet of Things—Enablers, Challenges and Applications

Communication Papers of the Federated Conference on Computer Science and Information Systems pp. 57–63
DOI: 10.15439/2019F87
ISSN 2300-5963 ACSIS, Vol. 20

A Framework for Network Intrusion Detection using Network Programmability and Data Stream Clustering Machine Learning Algorithms

Admilson de Ribamar Lima Ribeiro[1] , Edward David Moreno Ordonez[2] and Anderson Clayton Alves Nascimento[3]
[1,2] *Computing Department, Federal University of Sergipe, UFS*
São Cristóvão, Brazil
email: admilson@ufs.br, edwdavid@gmail.com
[3]*Institute of Technology, University of Washington Tacoma*
Tacoma, Washington
email: andclay@uw.edu

***Abstract*— Several operational security mechanisms have been developed to mitigate malicious activity in the Internet. However, the most these mechanisms require a signature basis and present the inability to predict new malicious activity. Other anomaly-based mechanisms are inefficient due to the possibility of an attacker simulates legitimate traffic, which causes many false alarms. Thus, to overcome that problem, in this paper we present an anomaly-based framework that uses network programmability and machine learning algorithms over continuous data stream. Our approach overcomes the main challenges that occur when develop an anomaly-based system using machine learning techniques. We have done an experimental evaluation to demonstrate the feasibility of the proposed framework. In the experiments, we use a DDoS attack as network intrusion and we show that the technique attains an Accuracy of 98.98%, a Recall of 60%, a Precision of 60% and an FPR of 0.48% for 1% DDoS attack on the real normal traffic. This shows the effectiveness of our technique.**

***Index Terms*--Operational Security, DDoS, Machine Learning, Data Stream.**

I. Introduction

THE HUGE variety of attacks in the Internet combined with the emergence of the new environments such as smart homes has demanded the improvement of forms of defense. This issue is more serious by fact that traditional mechanism of security as cryptography is not adequate to be used due to the real-time nature of these new environments. That is, very strong cryptograph functions can slow down the system. Thus, detection intrusion mechanisms are more adequate to face this issue.

Several studies have analyzed the use of detection intrusion in computer network security [1-7]. However, all those solutions are based in signature, that is, there is a need of a signature basis of the intrusions so that they can be detected. Despite its speed in detecting certain attacks due to the knowledge of the signatures of the main attacks, this category of solutions presents a crucial limitation: they can only detect attacks that are compatible with previously available signatures, not acting in new ways of malicious code.

One way to bypass that limitation, it is addressing the problem through a solution based in anomaly where abnormal traffic can be detected considering the knowledge of normal activities through profiles. Thus, deviations from normality are treated as threats. This technique presents two main problems [8]: a high false alarm due to the possibility of an attacker simulates legitimate activities and normally, it is used synthetical data due to difficult to find real training datasets. Moreover, some studies have employed machine learning techniques in inappropriate manner [9].

Therefore, in this paper, we analyze the challenges that must be overcome to provide an anomaly-based network intrusion detection mechanism and we present a framework that uses network programmability through SDN (Software Defined Networking) and a machine learning algorithm that works with continuous data stream clustering. In addition, the algorithm provides an outlier detection mechanism inside it. In our experiment, we use a DDoS attack as network intrusion to demonstrate the viability of the proposed framework

The remainder of this paper is organized into six sections. In Section II, we present the related works and highlights our contribution. In Section III, we present the main challenges in the development of an anomaly-based mechanism using machine learning algorithms. In Section IV, we outline a set of requirements to improve such mechanisms. In Section V, we describe our framework to meet the requirements presented in the Section V. In Section VI, we present and discuss some results. Finally, in Section VII, we discuss the future research and conclude the paper.

II. Related Work

Our anomaly-based solution mainly comprises three concepts of machine learning algorithms: clustering, outlier detection and data streaming so that we revised the literature considering those three factors.

In [10], the authors present an unsupervised solution based on a modification of outlier detection mechanism of random forest algorithm. The experiments were performed using KDD'99 dataset. The results were like results of previous approaches.

©2019, PTI

Devarakonda *et al*. [11] explore the possibility to detect outliers using a multi layered framework. This solution is adequate to high dimensional datasets. The experiments are performed using the KDD'99 dataset.

In [12], the authors present SPOT (Stream Projected Outlier Detector) to find out outliers in Unix system. SPOT can process high-dimensional data streams and detect new attacks. This solution use UNM datasets in the experiments.

Da *et al.* [13] present a method to detect DDoS attacks and SYN flood attacks. The method is an improved mining outlier detection using clustering. The experiments were executed in real time in a local network.

In [14], the authors combine several machine learning techniques to obtain an effective intrusion detection mechanism. The techniques were PCA, K-means and SVM. In the experiment was used KDD´99 dataset.

In [15] and [16], the authors present a self-protection architecture for IoT based on artificial neural network algorithms and fuzzy logic. The main DDoS attacks that occur in IoT environment has been investigated such as, selective forward, blackhole, sinkhole and flooding.

However, those studies that use machine learning algorithms are limited by using of static data usually located in datasets. Those algorithms use small datasets of training data that are available in memory. In some environments, such as IoT applications and TCP/IP traffic that generate high-speed data streams, it is impracticable to store all the data in memory and run multiple passes over the training data. Moreover, the learning model can change over time due to the generation of data by non-stationary distributions, for example the occurrence of a new malicious activity in the computer network. This change has an impact on the algorithm accuracy since the training dataset is soon outdated and there are errors in the estimator forecast. In addition, some systems like SDN networking, the switches do not select their ports and paths through static data, but through data streams. Therefore, the best way to analyze the collected information is not through datasets but through transient data flows. Thus, there is necessity of a new approach that considers transient data flows in the Internet. Our contribution in this paper is to show a framework to fill this gap.

III. Challenges in Anomaly-based Intrusion Detection System

Intrusion Detection System (IDS) is an operational mechanism to detect malicious activity in a computer network. These malicious activities or intrusions can affect the normal operation of system resulting in serious problems of security. As the malicious activities are different of normal behavior of system, anomaly detection techniques can be applied in this case [17]. Some techniques commonly used are machine learning models.

However, it is necessary to consider a set of challenges before to implement an anomaly-based mechanism using machine learning algorithms. We group these challenges in three categories: 1) according to the characteristics of network traffic, 2) according to the problems with anomaly detection approaches and 3) according to the application of machine learning techniques in inappropriate manner.

In the category 1, there are some key challenges in computer network domain including huge volume of data, data streaming fashion and high data variability. Nowadays, the data acquisition is automatic instead of manual. Sensor devices and computers collect, process and send information to other computers continually all the time resulting in huge volume of data. In some networks, such as wireless sensor networks and TCP/IP networks, the best manner to collect and analyze the data is through continuous data stream instead of persistent files. Thus, techniques based on small datasets and in batch are not more viable. Moreover, normal traffic in data stream present high variability, becoming difficult to find stable patterns.

In the category 2, anomaly detection approaches present the following problems: high false alarm rate, training data lack and evasion [8]. The detectors normally generate a high false positive rate, and in network domain that is more problematic due to the huge volume of data. A determined rate is more significant than other domains with a volume of data reduced. In relation a training data, it is not difficult to find labeled data for normal behavior, but it is hard to find labeled data for the intrusions. Finally, it possible for an attacker simulates malicious activity as benign activity so that the system can be evaded.

In the category 3, the use of machine learning techniques in anomaly detection presents the following challenges: outlier detection, high cost of errors, semantic gap and difficulty of evaluation [9]. Although there is a necessity in detecting outliers, machine learning algorithms present higher performance in detecting similarities instead of outliers. That is, that technique is more adequate to detect similarities among the data. In classification problems, the error rate is high when compared with other domains, sometimes damaging all the system. Not always, the detection results are interpreted correctly by the network operator, occurring a semantic gap that need to be resolved. Finally, the difficulty of evaluation is a significant challenge due to the lack of real data to work in this domain.

IV. Proposed Overcomings

A. Overcoming the Category 1 Challenge

For facing this challenge, we use a data stream machine learning algorithm. In machine learning system that apply continuous data stream, it is generated a non-stationary and dynamical environment, which the data are achieved in a continued way and with dynamical unknown. The learning process is continuous and evolves over time. Machine learning algorithms can incorporate new information into the decision model, detect and react to changes, using limited computational resources. Changes that evolve over time include: the user's interest, the type of anomaly, the quality of a product, among others [18].

The properties of a computational model in these learning systems are [18]: 1) incrementality; 2) real-time learning; 3)

to be able to process examples in constant time and limited memory; 4) limited access to processed examples; 5) capacity to detect and to adapt the decision model to concept drift.

The properties from 1) to 4) are adequate to apply in the computer network domain because of the huge volume of data, data streaming fashion and high data variability. Due to the huge volume of data, it is adequate that anomaly-based techniques are efficient computationally to handle that challenge. The characteristic of data streaming demand an incremental and online approach. The property 5) is adequate to face the high data variability due to adaptation of model for new instances.

B. Overcoming the Category 2 Challenge

The machine learning technique is not a panacea that can be employed in all anomaly problem successfully. In this type of problem, it is necessary to have a clear view about the objectives to be reached. The easier manner to get this view is maintain the objectives well limited. That is, try keeping the scope narrow to adapt the detector the specifics of problem and reduce the potential of errors of misclassification [9]. In our problem, for keeping the scope narrow, we treat just one type of attack: the DDoS attack. This attack consists in to try impeding that legitimate users access the services of a system, that is, becoming unavailable the resources or services provide for a computer network. This is done through the exhaustion of system resources such as servers, communication channel, etc. Normally, this attack is executed of a distributed (Distributed Denial-of-Service – DDoS) manner where a botnet is created to generate the attack. Thus, an attacker boosts a DDoS attack through a computer network instead of a simple computer. Thus, we adapt the detector to specifics of this attack and reduce the errors of misclassification.

The data availability is a crucial issue in anomaly detection problem. Working with real data is important because shows that system can used in practice. In our case, we overcome this challenge through using real data of an enterprise computer network.

Evasion occurs when an attacker can simulate malicious traffic as normal traffic so that the system is deluded. For reducing this threat, we consider an SDN environment with a small network that can be used to protect a larger network in an enterprise environment. The site presents low risk for explicit targeting for an attacker and, we consider that there will be a few evasion problems.

C. Overcoming the Category 3 Challenge

Machine learning techniques applied to anomaly detection problems are not well succeeded in relation to other domains. According to Sommer and Paxson [9], the difficulties that appears are due to use of machine learning in an inappropriate manner and point out some recommendations, that we follow, to bypass this problem: understanding the threat model, keeping the scope narrow, reducing the costs and treating evaluation issues.

The high rate of false positive is the principal factor to increase the costs in anomaly detection system. To overcome this problem, we follow three recommendations: 1) reduction of scope of system; 2) dealing with the traffic diversity through the machine learning algorithm over continuous data stream; and 3) rigorous examination of the features of network traffic.

In anomaly detection problem, the semantic gap can appear in two manners: in network operator actions and in the type of networks. Network operators can have difficulties in act in face of determined results. Normally, it is hard to identify what occurring in the network. That is, there is an anomalous traffic or there is a false alarm? We face this challenge through employing a mitigation technique of DDoS attacks, releasing the network operator of this problem. In relation to type of networks, academic networks have security policies different of corporate networks. That is, an anomaly detection system for corporate networks is not adequate for academic networks. Here, we are considering a corporate network.

About evaluation issues, we analyzed manually the false positive rate to check out if there is a case related incorrectly. And, we analyzed the true positive rate and true negative rate to verify if the system learned what it must learn.

In relation to the outlier detection, it is better to use an unsupervised technique. Unsupervised techniques of machine learning are more adequate to true classification problems, therefore with anomaly detection is better used to find out variations of known attacks instead of attacks themselves [17]. Thus, we can train the system with a few known attacks along with normal traffic available, thus overcoming this challenge.

V. FRAMEWORK OVERVIEW

Our framework is constituted of three components: SDN architecture, data stream clustering algorithms and mitigation technique. These components were chosen to implement the recommendations presented in the section IV.

A. SDN Architecture

Nowadays, the most of computer networks that belongs the Internet present an architecture where the control plane and the data plane are highly coupled and embedded in the same network devices. The whole structure is decentralized. This fact was very important in the early days of Internet where the main preoccupation was based on its resilience. That architecture is relatively static and very complex occurring management and innovation problems.

To overcome those problems, two proposals have appeared: to support network management, a small number of vendors offer proprietary specialized hardware solutions, operating systems and control programs; and to overcome the lack of new internal functionality, many specialized components have proliferated in today´s networks, such as firewalls, packet filters, packet inspection machines and so

on. These proposals have increased the complexity of management and innovation of the computer networks and their operations.

Due to the limitations of those proposals, a new approach of network architecture has emerged: the SDN architecture. This architecture is based on four pillars [19]: the control and data plane are decoupled; forwarding decisions are flow-based instead of destination-based; control logic is moved to external entity called SDN controller; and the computer network is programmable through software applications. Fig. 1 shows an SDN architecture with all the main components. The controller is responsible to provide abstractions and essential resources for facilitating the programming of forwarding devices. The forwarding devices execute an elemental operations set to forward network packets to hosts and other network elements.

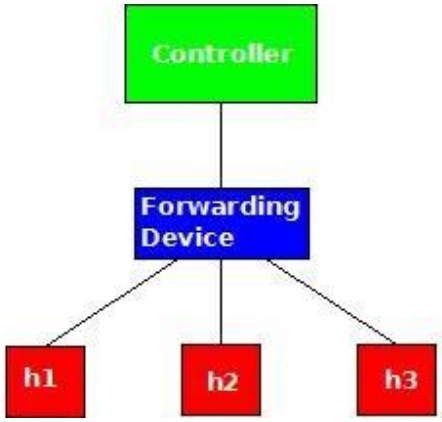


Fig.1 The SDN architecture.

The SDN networks present various advantages in relation to the traditional networks [20]: 1) It is easier to program network applications because now there are several abstractions provided to programming languages available. 2) The integration of different applications becomes more direct, for example it is possible combine a load balancing application with a routing application. 3) the network applications can take actions in any part of computer network. 4) the weak coupling between the control and data plane becomes easier to add new functionality.

We use SDN in our framework because of the easiness of implementing the machine learning algorithm as a network application rather than inserting some hardware device (middlebox) as would be done in traditional networks.

B. *Data Stream Clustering Algorithms*

Researches in data stream clustering algorithms have produced several proposals to provide unsupervised learning. Silva *et al.* [21] describe the main characteristics these algorithms considering some criteria such as window model, outlier detection mechanisms, cluster shape and so on. Considering this study and our problem where an algorithm of machine learning has to present three properties: clustering, outlier detection and data stream. The selected option was the OutlierDenStream algorithm [22].

This algorithm is a modified version of DenStream algorithm [23] and it is adequate to be used in unsupervised environment and in outlier detection. It uses the same concepts than DenStream algorithm: micro-cluster, distances, weight, neighborhood and pruning strategies. These concepts are represented in parameters. Therefore, OutlierDenStream uses a number of parameters (namely, μ, β, λ, and ε). The most of parameters are tuned automatically from OutlierDenStream, except the parameters λ and β that were tuned manually before evaluating the solution. The parameter λ represents the fading factor, that is, it is a weight to eliminate the core cluster before later instances can be added. The parameter β is related to pruning strategies.

The OutlierDenStream run in two phases. In the first phase is necessary to form the clusters and this is done through the DBScan algorithm [24] that builds a buffer dataset. After that, the algorithm maintains the clusters incrementally and when arrives a new sample it labels it as normal or abnormal, that is, the algorithm attempts to cluster normal instances, treating outliers as anomalous traffic.

C. *Mitigation Technique*

For mitigating the DDoS attack, we can employ the characteristics of the flow table of the SDNs. One manner is through the blocking of an entry of the table that supports a link where there is a DDoS attack. It is possible to block that link for a time fixed and after this time the link can be unblocked. Thus, users will not disturb by the activity of security of the computer network.

In order to show how the mitigation technique works, we can consider the scenario shown in Fig. 2.

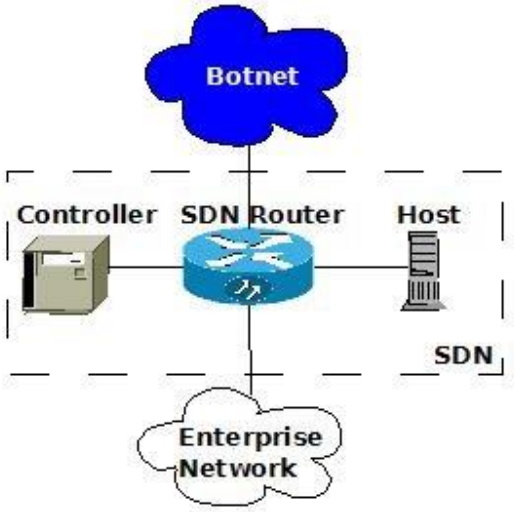


Fig. 2 The SDN Scenario.

The botnet created for an attacker try to attack an enterprise network through an DDoS attack. The enterprise network has access to the Internet through a SDN router that contains a flow table to forward the network packets since authorized by controller. Our data stream clustering algorithm is deployed in the controller because it is responsible to verify all the ingoing traffic. The normal traffic is observed by a certain period of time. Afterwards, the data stream clustering algorithm acts to verify if there is some abnormality in the network. If an DDoS attack is detected, an action is taken in the controller to block the entry in the flow table that supports the suspect traffic. After, a certain time this entry in the flow table is unblocked.

The viability of this technique is possible due to two factors: only one entry in the flow table is blocked and quickly released and also the short duration of DDoS attacks. Thus, it is possible that the detection algorithm can be executed in the controller.

VI. EXPERIMENTAL EVALUATION

In our experiment, we implemented the anomaly-based mechanism using a DDoS attacks as anomalous traffic and we simulated the data streaming through a real dataset of a corporation computer network.

The algorithms were obtained through Python libraries at the 3.5.1 version. The execution of the algorithms to get the best hyperparameters was in a HP desktop, with 4Gb of RAM, Intel i5-3470S (2.9GHz) executing in an Ubuntu 12.04.5 LTS operational system.

A. Data Stream and Processing

The used dataset contains traffic of a corporation network that was attacked for a period of 24 hours [25]. It was selected 76 features as shows the Table I. The normal behaviour contains traffic of five different activities such as checking of e-mail and file transferring among the users of the company. The normal traffic was labelled as “Bening” and the DDoS attack traffic was labelled as “Attack”.

For the experiments, we use four datasets with 1%, 2%, 3% and 4% of DDoS attack traffic. Each dataset was generated of a unique dataset that contains 464976 labelled instances of training and 4% of DDoS attack traffic. Table I shows the description of each dataset.

Table I. Dataset Description

Number of dataset	Number of features	Percentage of DdoS attack traffic	No of labelled instances
1	76	4 %	464976
2	76	3 %	461464
3	76	2 %	456814
4	76	1 %	452163

For evaluating our technique, the following performance metrics were used: Accuracy, Precision, Recall and False Positive Rate. The results of anomaly detection are commonly represented in a confusion matrix composed of TP (True Positives), FN (False Negatives), TN (True Negatives), and FP (False Positives), respectively. The Precision and Recall are defined as: Precision=TP/(TP+FP), Recall=TP/(TP+FN). The Accuracy is defined as: Accuracy = NCD/TI, where NCD is the Number of Correct Detections and TI is the Total of Instances. The False Positive Rate is defined as: FPR = FP/(TN+FP).

The Accuracy metric shows the capacity of success of the technique, the Precision metric measures how well the technique detected abnormal instances, the Recall metric complements the Precision metric for all the instances, and the False Positive Rate (FPR) indicates the percentage of false alarms generated, in this case, normal traffic identified as DDoS attack.

It is common in anomaly detection problem to present the results as ROC (Receiving Operating Characteristics) curve [26]. The ROC curve plots the detection rate, represented by the Recall metric, against the False Positive Rate. Thus, it is possible evaluate the performance of an anomaly detector through relating to two performance metrics.

B. Results and Discussion

After the selection of performance metrics, we use the dataset of an enterprise computer network. The instances are converted in data stream by taking the data input order as the order of streaming. In the data streaming, appears only two types of network traffic: normal traffic and DDoS attack traffic.

For adequate use of the OutlierDenStream algorithm, we must tune, manually, the parameters λ and β using a set of training data. For simulating the data stream, we collect 452163 instances of normal traffic and 1% of DDoS attack traffic. Fig. 3 shows the variation of performance metrics (Recall, Precision, Accuracy and FPR) against the parameter λ. We can observe the best choices are in the interval from 0.015 to value 0.03. The value 0.03 was used in our validation. The parameter β have not presented any variation in the metric performances and was established to zero in all experiments.

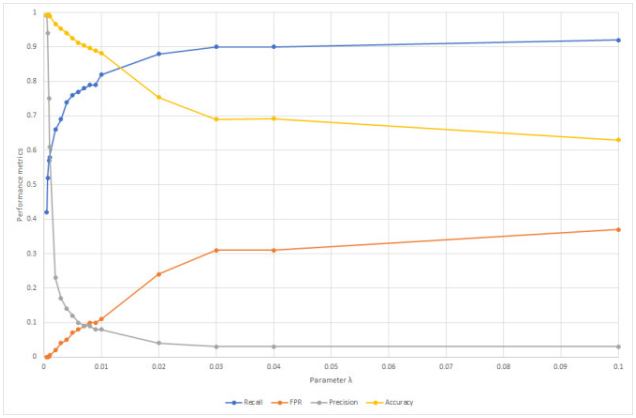


Fig. 3 The performance metrics versus parameter λ. – dataset 4

After 452163 instances of data stream, the OutlierDenStream algorithm achieved an Accuracy of 98.98%, Precision of 60%, Recall of 60% and FPR of 0.48%. It is important to point out the low FPR, because, when a false positive is identified in anomaly problem, a normal traffic can be blocked as a mitigation action. A false positive requires expensive time of a network administrator to examine a problem that did not occur or processing time if the mitigation is made automatically.

Fig. 4 shows the ROC curve for the dataset that contains 1 % of DDoS attack (dataset 4). We can observe that our technique can achieve a high detection rate for a low false positive rate. We can note that the best combination is a 0.9 Recall and a 0.2 FPR.

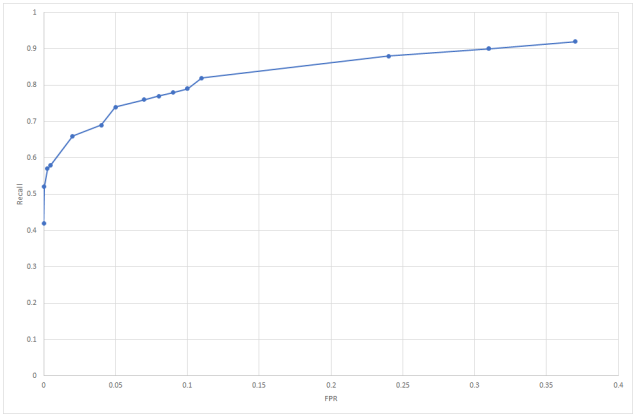


Fig. 4 The ROC Curve – dataset 4

Fig. 5 shows the Precision-Recall plot for the dataset 4. We can observe that the best combination is a 0.6 Precision and a 0.98 Recall. We can see that it is possible to improve the Precision but, in this case, we have a high decrease in Recall.

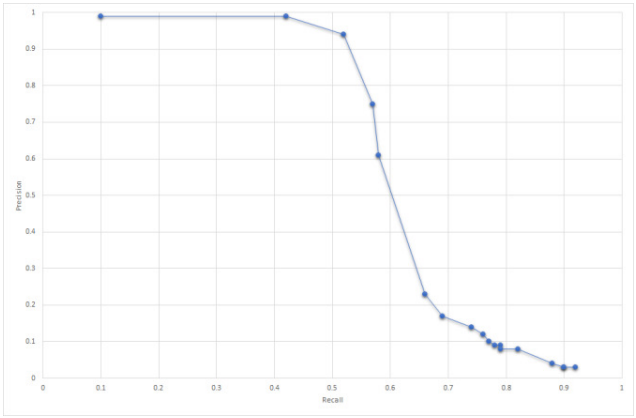


Fig. 5 The Precision-Recall plot – dataset 4

The most of machine learning algorithms present overfitting. Overfitting occurs when there are a lot of errors on instances data that has not been trained, that is, the machine learning algorithms cannot generalize. Basically, there are two manners to prevent overfitting: using the cross-validation technique or split instances in different datasets (training and testing datasets). In our experiments, we use four different datasets and measure the difference in Recall between them. As shown in Fig. 6, the difference between the Recall is less than 2%, that confirm that there is not overfitting on the algorithm used.

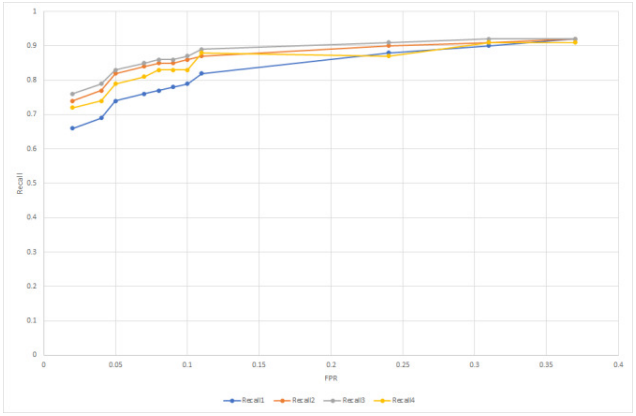


Fig. 6 The recall metric – datasets 1,2,3 and 4

VII. Conclusions and Future Work

In this paper, we have used an SDN architecture, data stream clustering machine learning algorithms and a mitigation technique to design a security framework for detecting and mitigating the DDoS attacks that can occur in the Internet. With this framework will be possible to protect any corporate computer network connected to the Internet.

From experiments, we can observe the effectiveness of our solution through the performance metrics used. On normal traffic, the 1% DDoS attack attains an Accuracy of 98.98%, a Precision of 60%, a Recall of 60% and FPR of 0.48%, while the 4% DDoS attack attains closer values of

the 1% DDoS attack attains an Accuracy of 98.98%, a Precision of 60%, a Recall of 60% and FPR of 0.48%, while the 4% DDoS attack attains closer values of the 1% DDoS attack. We can also note that in our framework it is possible to improve the precision decreasing the recall.

In the future works, we will implement our solution in an online and real environment. Besides, we will implement the mitigation technique of DDoS attacks inside the SDN environment. Also, we will develop new techniques to avoid various attacks that can occur in the Internet using the SDN environment to protect a computer network.

REFERENCES

[1] B. Sun, L. Osborne, Y. Xiao, et al. Intrusion detection techniques in mobile ad hoc and wireless sensor networks. IEEE Wirel Commun 2007; 14: 56–63.

[2] V. Paxson. Bro: a system for detecting network intruders in real-time. Comput Netw 1999; 31(23): 2435–2463.

[3] J.B. Cabrera, L. Lewis, X. Qin et al. Proactive detection of distributed denial of service attacks using MIB traffic variables—a feasibility study. In: Proceedings of the 2001 IEEE/IFIP international symposium on integrated network management, Seattle, WA, 14–18 May 2001, pp.609–622. New York: IEEE.

[4] M. Roesch. Snort: lightweight intrusion detection for networks. In: Proceedings of the 13th USENIX conference on system administration, Seattle, WA, 7–12 November 1999. Berkeley, CA: USENIX Association.

[5] C.M. Cheng, H. Kung and K.S. Tan. Use of spectral analysis in defense against DoS attacks. In: Proceedings of the IEEE global telecommunications conference, 2002 (GLOBECOM' 02), Taipei, Taiwan, 17–21 November 2002, vol. 3, pp.2143–2148. New York: IEEE.

[6] A. Hussain, J. Heidemann and C. Papadopoulos. A framework for classifying denial of service attacks. In: Proceedings of the 2003 conference on applications, technologies, architectures, and protocols for computer communications, Karlsruhe, 25–29 August 2003, pp.99–110. New York: ACM.

[7] A. Hussain, J. Heidemann and C. Papadopoulos. Identification of repeated denial of service attacks. In: Proceedings of the 25th IEEE international conference on computer communications (INFOCOM 2006), Barcelona, 23–29 April 2006, pp.1–15. New York: IEEE.

[8] C. Gates and C. Taylor, "Challenging the Anomaly Detection Paradigm: A Provocative Discussion," in Proc: Workshop on New Security Paradigms, 2007.

[9] R. Sommer and V. Paxson, "Outside the Closed World: On Using Machine Learning for Network Intrusion Detection", In Proc. of IEEE Symposium on Security and Privacy, pp. 305-316, 2010.

[10] J. Zhang and M. Zulkernine, "Anomaly based network intrusion detection with unsupervised outlier detection," in *Proc. 2006 IEEE International Conference on Communications (ICC)*, 2006, vol. 5, pp. 2388-2393.

[11] N. Devarakonda, S. Pamidi, V. V. Kumari, and A. Govardhan,"Outliers Detection as Network Intrusion Detection System Using Multi Layered Framework," in *Advances in Computer Science and Information Technology*, Communications in Computer and Information Science Vol. 131, 2011, pp. 101-111.

[12] D. Kershaw, Q. Gao, and H. Wang, "Anomaly-based network intrusion detection using outlier subspace analysis: A case study," in *Advances in Artificial Intelligence*, Lecture Notes in Computer Science Vol. 6657, 2011, pp. 234-239.

[13] W. Da and H. S. Ting, "Distributed intrusion detection based on outlier mining," in *Proc. 2012 International Conference on Communication, Electronics and Automation Engineering (ICCEAE)*, Advances in Intelligent Systems and Computing Vol. 181, 2013, pp. 343-348.

[14] P. Manandhar, and Z. Aung. "Intrusion Detection Based on Outlier Detection Method". Intl' conference on Intelligent Systems, Data Mining and Information Technology (ICIDIT'2014) April 21-22, 2014 Bangkok (Thailand)

[15] R. M. A. C. Mello; A. R. L. Ribeiro; F. M. Almeida, and E. D. Moreno."Mitigating attacks in the Internet of Things with a Self-protecting Architecture". In: AICT 2017 - The 30th Advanced International Conference on elecommunications, 2017, Venice. Proc. of AICT 2017. Paris: IARIA, v.1. p. 1-6, 2017.

[16] F. M. Almeida; A. R. L. Ribeiro; E. D. Moreno; and C. A. E. Montesco. . "Performance Evaluation of an Artificial Neural Network Multilayer Perceptron with Limited Weights for Detecting Denial of Service Attack on Internet of Things". In: In: AICT 2016 - The 12th Advanced International Conference on Telecommunications, Valencia. Proc. of AICT 2016. Paris, France: IARIA XPS Press, v. 1. p. 1-6, 2016.

[17] V. Chandola, A. Banerjee and V. Kumar. "Anomaly Detection: A Survey", ACM Computing Surveys, pp 1-72, 2009.

[18] P. Domingos and G. Hulten. A general method for scaling up machine learning algorithms and its application to clustering. In: Proceedings of the Eighteenth International Conference on Machine Learning, p. 106-113, 2001.

[19] J. A. Wickboldt, W. P. de Jesus, P. H. Isolani, C. B. Both, J. Rochol, and L. Z. Granville, "Software-Defined Networking: Management Requirements and Challenges," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 278–285, Jan 2015.

[20] D. Kreutz, P. E. Veríssimo, S. Azodolmolky, "Software-Defined Networking: A Comprehensive Survey", arXiv preprint arXiv:1406.0440, 2014

[21] J. A. Silva, E. R. Faria, R. C. Barros, E. R. Hruschka, A. C. P. L. F. Carvalho and J. Gama, "Data Stream Clustering: A survey", ACM Computing Surveys, vol 46, Issue 1, October 2013.

[22] "The OutlierDenStream Algorithm", https://github.com/anrputina/OutlierDenStream.

[23] F. Cao, M. Ester, W. Qian and A. Zhou, "Density-based Clustering over an Evolving Data Stream *SIAM Conference Data Mining*, Bethesda, 2006.

[24] M. Ester, H. Kriegel, J. Sander and X. Xu, "A Den-sity-Based Algorithm for Discovering Clusters Spatial Databases with Noise," *International Conference on Knowledge Discovery in Databases and Data Mining* (*KDD*-96), Portland, pp. 226-231, 1996.

[25] A. Ghorbani and A. H. Lashkari, CDMC2018 Dataset: DDoS Attacks Detection for Enterprise Network Security, Canadian Institute for Cybersecurity, University of New Brunswick, http://www.csmining.org/

[26] T. Fawcet, "An introduction to ROC analysis", Pattern Recognition Letters, p. 861-874, 2005.

3rd Workshop on Internet of Things—Enablers, Challenges and Applications

THE 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,** Univiversita 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
- **Hudik, 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 Telecomunications 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, Niranjan,** Institute of Human and Machine Cognition
- **Teslyuk, Vasyl,** Lviv Polytechnic National University, Ukraine
- **Wang, Zhonglei,** Karlsruhe Intitute 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

Communication Papers of the Federated Conference on Computer Science and Information Systems pp. 67–72
DOI: 10.15439/2019F293
ISSN 2300-5963 ACSIS, Vol. 20

Solutions for Planning Smart Hybrid Public Transportation System – Poznan Agglomeration as a Case Study of Satellite Towns' Connections

Marzena Banach[1] Katarzyna Kubiak[2] and Rafał Długosz[3]
[1] Poznan University of Technology, Institute of Architecture and Spatial Planning,
ul. Nieszawska 13C, 61-021 Poznan, Poland
E-mail: marzena.banach@erba.com
[2] Adam Mickiewicz University
Faculty of Mathematics and Computer Science
ul. Umultowska 87, 61-614 Poznan, Poland
E-mail: kbkubiak45@gmail.com
[3] UTP University of Science and Technology
Faculty of Telecommunication, Computer Science and Electrical Engineering
ul. Kaliskiego 7, 85-796, Bydgoszcz, Poland
E-mail: rafal.dlugosz@aptiv.com, rafal.dlugosz@gmail.com

***Abstract*—A primary goal of the Internet of Things is to create smart spaces, including smart cities and intelligent transportation system (ITS). One of the ITS variants is smart hybrid public transportation system, whose efficient development requires a broad support from the side of information and communication technologies (ICT). For such a system to function optimally, vehicles that are part of it have to be able to exchange data with other road users and road infrastructure. This applies especially to autonomous vehicles, whose trajectories may be in the future controlled through the vehicle-to-infrastructure (V2I) communication network. In the paper we propose a system that can be used to optimize the boundaries of relatively small suburban zones, in which public autonomous vehicles would be allowed to operate. Vehicles of this type may offer the transport only to the nearest local railway stations, from which the travel to the central city of the agglomeration would be continued for example by train. The concept of the proposed solution is based on the Voronoi diagrams, in which particular suburban train stations are treated as local attractors. The proposed software system in one of its stages uses Google Maps engine that allows for the determination of road distances and travel times between particular towns. On the basis of such data it determines the mentioned zones. Their boundaries, as well as optimal routes in a given period of the day may be communicated to the vehicles through the V2I system. The system performance is presented for an exemplary case of Poznan city, Poland.**

***Keywords*—Intelligent Public Hybrid Transport System, Transport Planning, Autonomous Vehicles, Software System**

I. Introduction

The idea of smart cities is frequently connected with technical and social solutions which aim at improving the quality of broadly understood everyday functioning of city dwellers. These solutions include various facilities, in which novel artificial intelligence (AI) as well as information and communication technologies (ICT) are used as a basis [1]. For example, these technologies are used to facilitate the access to information, offering information systems for passengers, tourists, etc. On a higher level, more advanced solutions are proposed such as remote access to offices [2]. However, one of the main pillars of smart cities is the so-called intelligent transportation system (ITS) [3], which bases on various AI and ICT technologies to a high extent. The ITS is expected to transform the transport toward more functional, flexible and ecological solutions. A truly intelligent system of this type should offer comprehensive solutions at the level of larger areas such as city agglomerations. This aim can be accomplished by developing an intelligent public hybrid transportation system.

Technological solutions used in smart cities should primarily lead to real social benefits in the areas of safety, ecology and economy. With appropriate approach, these benefits can be directly offered by the ITS. One of them, which is the topic of this work, is to improve the competitiveness and flexibility of public transport in towns and villages – satellites of major cities in agglomerations. Their inhabitants, who constitute a significant percentage of all inhabitants of the agglomeration, regularly commute to the central city, mostly for educational and occupational purposes. The lack of an attractive public transport offer, e.g. rare and inconvenient connections, long travel time, causes the car to be a frequently used means of transport (in Poland). It has lead to a growing concern in recent years, which is increasing city traffic which adversely affects the quality of life.

Not unlike most areas of life, transport is subject to market competition. In terms of public transport, a competitive offer is such an offer which is comparable to what individual transport provides. Simply speaking, people demand faster, cheaper, more convenient and more flexible solutions.

One of the solutions to the described problem is to use a combination of two means of transport: own car and public transit. This idea is not new. It has been promoted for many

©2019, PTI

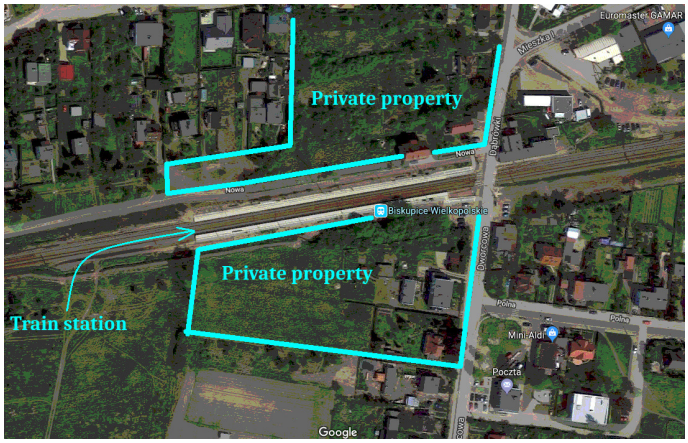


Fig. 1. The environment of the Biskupice Wielkopolskie train station – 20 km from Poznan city, Poland.

Fig. 2. Public hybrid transport system in Poznan [4].

years in numerous agglomerations around the world. Usually, parking lots are created close to a train station and are available to people who change to a train. However, it is not always possible in practice. For example, there is no room for a sufficiently large parking lot at many railway stations in Poland. Figure 1 presents the railway station in Biskupice Wielkopolskie, a village located about 20 km from Poznan (Poznan is the central city of the agglomeration). It is not possible to build a parking lot at this station due to the lack of space. Biskupice Wielkopolskie and neighboring villages are inhabited by more than 2,000 people, many of whom travel to Poznan almost every day. This means even more than hundred additional cars on the roads of Poznan. It is a relatively common problem in Poland which has to be solved.

In recent years, the automotive industry has been dynamically developing. One of the hot directions is to popularize autonomous vehicles, developed by many car manufacturers. The aim of such solutions is primarily to improve road safety through the elimination of mistakes made by drivers. Another important goal is to optimize the driving style, which will lead to the reduction in pollutant emissions.

However, it is worth to pay attention to a few basic aspects related to autonomous vehicles. One of them is the large complexity of the environment "seen" by such vehicles. In cities, for example, it means a large number of objects (pedestrians, other vehicles, bicycles) in the vicinity of the car which have to be tracked by its systems. Another problem can be a long transition period, during which autonomous vehicles will coexist with conventional cars, which raises safety issues. In spite of these problems, the emergence of autonomous vehicles may have positive impact on public hybrid transport. With some assumptions, such cars can be used in the described transition period. We consider using such cars in limited, well defined and flexible as well as less dense suburban areas.

In recent years, much attention has been paid to public hybrid transport, introduced in many cities. The aim is to achieve such flexibility of this transport that possibly many residents will be encouraged to give up using own cars. A joint tariff and joint schedule plans take into account buses, trams, subway (if available), urban and suburban trains. In urban literature, the problem of the so-called first and last kilometer is highlighted. According to this rule, if reaching the nearest public transport stop consumes too much time (over 10 minutes/1 km), the residents usually choose their own car. New concepts of compact cities with buildings located near to public transport stops will satisfy most of the needs of the inhabitants, so that it would not be necessary to cover long distances frequently. One of the basic goals of this concept is to eliminate the need to use own car on a daily basis.

The solutions discussed in this work are not aimed at polemics with the idea of compact cities postulated by modern urbanism. The main goal is to facilitate the access to the city resources, including the case of wider agglomerations. Optimization of public transportation using autonomous vehicles (taxis) outside central cities can shorten the access time to the city resources, while maintaining ecological goals. One can talk here about an apparent shortening of the distances.

As mentioned above, the idea of the hybrid public transportation system is known and developed for many years over the world. New trends in this area, however, are various concepts of the use of autonomous vehicles in public transport. One of the solutions here are autonomous taxis already used in some cities (e.g. NuTonomy in Singapore). In the scope of our interests, on the other hand, are technical solutions that aim at optimizing public transport away from the major cities of the agglomerations [5]. We think this will have a positive impact on the overall area of particular agglomerations. Such optimization can be understood, for example, as minimizing the total distances traveled by the inhabitants of particular suburban areas by private cars. At the same time, the goal is to keep the flexibility offered currently by own means of transport in these areas to the highest possible level $\rightarrow$ minimizing the total travel time. In this work, we deal with technical solutions for a system that can be used for planning the hybrid transport in suburban area. In particular, we focus on defining boundaries of particular zones available for selected groups of autonomous vehicles. To illustrate the complexity of the problem, it is worth mentioning that there are over 100,000 cities and towns in Poland that need to be included in planning.

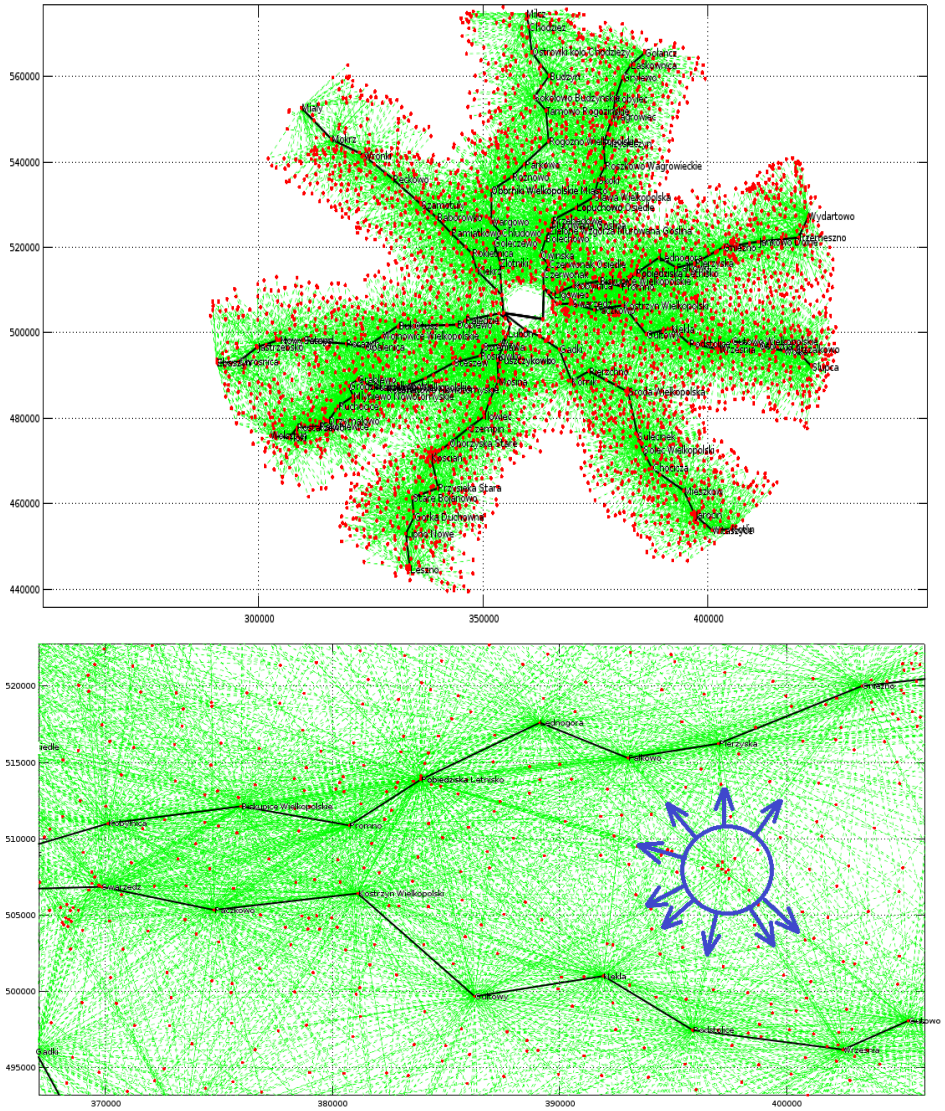


Fig. 3. Illustration of data stored in the $DT2_{l,s}$ datasets: selected settlement-to-station assignments (up). Exemplary area belonging to ROIs of nine stations for $r_{\max}$ set to 22 km (bottom).

The paper consists of several parts. The next section presents a short literature study related to the subject of the work. We briefly discuss public hybrid transport systems and the use of Voronoi diagrams for agglomerations. In the following section we describe the proposed software system, as well as exemplary results for the Poznan agglomeration in Poland. The conclusions are formulated in last section.

II. State-of-the art study and research methodology

The investigation results presented in this work are for an exemplary case of the Poznan agglomeration. However, the proposed software system can be quickly configured and used for any agglomeration. The methodology applied in our investigations partially results from the following state-of-the art study which showed us the areas of the application of the system.

A. Public hybrid transport systems in Poznan

Combining different means of transport is an increasingly common practice, especially in larger urban agglomerations. The common tariff plan connects buses, trams, subway and suburban trains wherever possible. This combination also includes the synchronization of timetables so as to improve the smoothness of getting around the city. An example of these solutions is the transport system in the Poznan agglomeration. In this case, the BTK (bus, tram, train) offer extends within a radius of approximately 50 km from Poznan, as shown in Fig. 2. The investigation results, presented in next section, were conducted for a similar area.

B. Using Voronoi diagrams

The proposed work is based on the method of dividing the space into ‘zones of influence’, which are often illustrated with the help of the Voronoi diagrams. This method is also

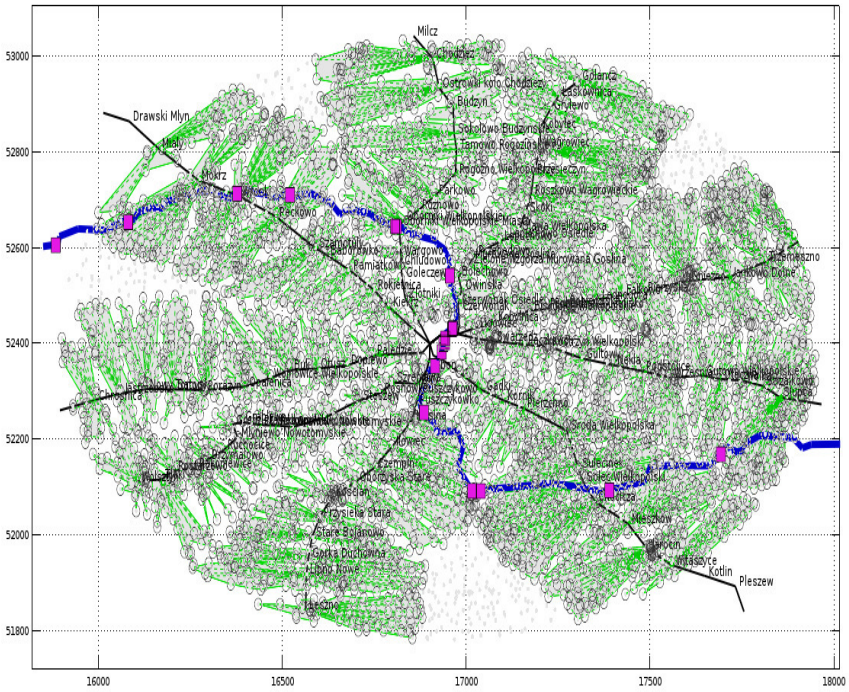


Fig. 4. An exemplary Voronoi diagram created by the proposed system of the overall Poznan agglomeration for the radius of 50 km.

used in urban planning [6] – [10]. An interesting example of using this methodology for planning the public transport is reported in [11]. In [11] a station location method for a commuter rail system has been demonstrated for an example case of Brasilia Metropolitan Area. This case is somehow similar to what we observe in Poland – traveling even large distances due to, for example, job opportunities in larger cities. The authors of [11] in their work focus on planning new rail routes. We investigate a different problem. We treat existing railway stations as reference points, while we are interested in optimizing access to particular stations from nearby towns.

C. Input data for the investigation

To make the investigations comprehensive and to increase the reliability of the tests of the implemented system, we used an initial dataset consisting of parameters of 120,000 Polish cities, towns and villages. Every entry provides the name of the unit, its county and its geographical location (longitude and latitude), among others. It was used to determine which settlements lie within a given radius from Poznan and will be denoted as DT0 from now on.

III. Proposed system and investigation results

We used the following notation to denote the datasets used in calculations: capital letters indicate the type of data stored, while the following number indicates the step of calculations. Also, each step has input and output data, as shown in Table I.

TABLE I
Inputs and outputs of particular computation steps (explanation of abbreviations in text).

Step No.	Inputs	Outputs
1	DT0	DT1
2	DT1 DR1 $r_{\min}$, $r_{\max}$	$DT2_{l,s}$
3	$DT2_{l,s}$	$DT3_{l,s}$
4	$DT3_{l,s}$ DT1	DV4
5	DV4	Voronoi diagrams

Step 1: Selection of units belonging to a given agglomeration

The aim of the first step is to identify the units from the input dataset DT0 whose straight line distance (calculated from their geographical coordinates) from Poznan is less or equal to 50 km. This value is based on the reach of the above-mentioned BTK transport system. This step was implemented in the C++ language. The resultant dataset is composed only of nearly 4,000 settlements located in the investigated area. It will be denoted as DT1 from now on.

Step 2: Building regions of interest for particular train stations

Firstly, we created a dataset, denoted as DR1, of subsequent railway stations in the investigated area and their geographical coordinates. They are grouped into 9 train lines leading to Poznan. In order to simplify the following computations, every line and every station within a line have their own identification numbers, denoted ad ID. Consequently, the DR1 dataset provides the so-called local central points/local attractors which gather road traffic from the nearest settlements.

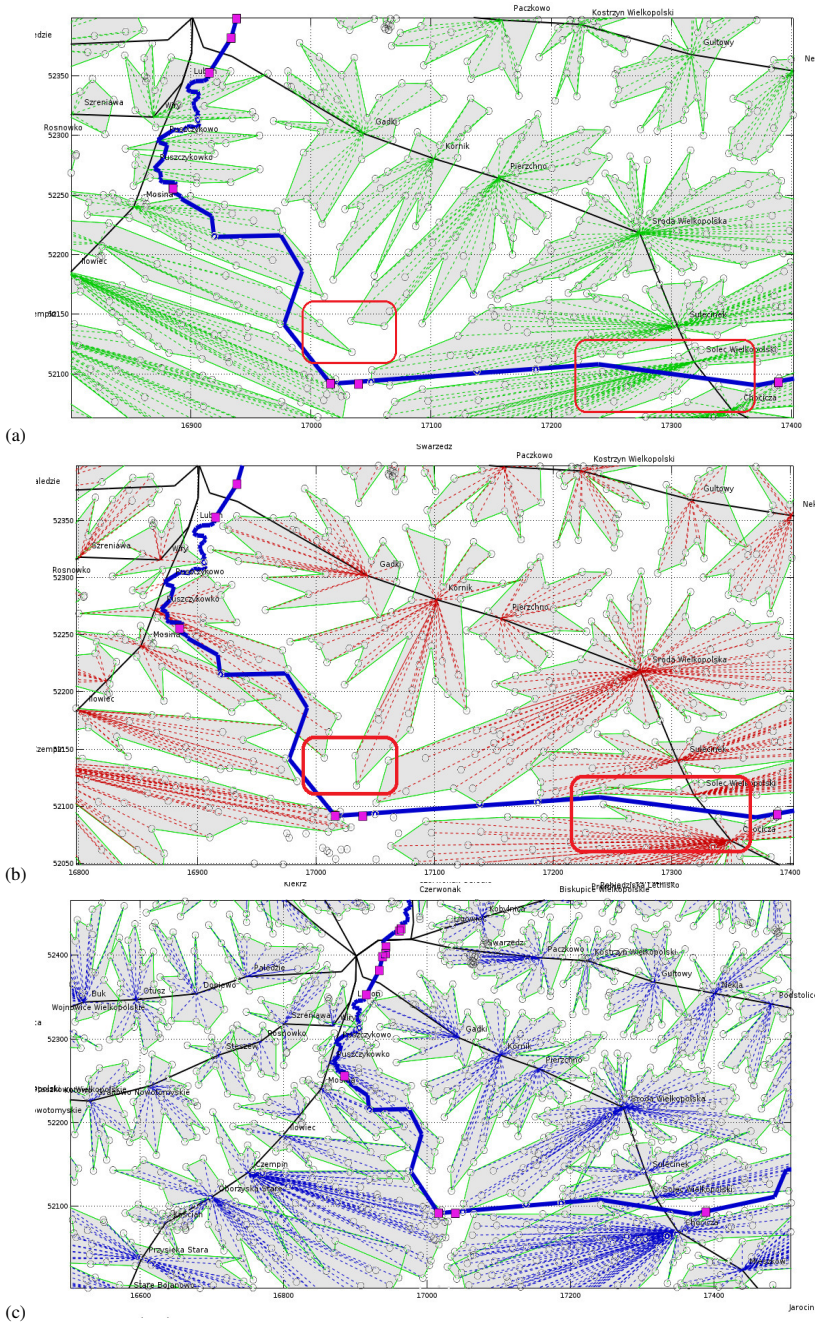


Fig. 5. Voronoi diagrams of selected areas for geographical (a), road (b), and travel time distances (c). The wide blue line is the Warta river, while the magenta blocks are locations of bridges

Then, we define a circular region of interest, denoted as ROI, for every railway station from DR1. ROI is described by two parameters the minimum radius $r_{\min}, r_{\min} > 0$ and the maximum radius $r_{\max}$. Selecting a non-zero value of the $r_{\min}$ is necessary to exclude internal districts of cities, in which particular train stations are located, from further computations.

Finally, for each station from DT1 a dataset $\mathrm{DT2}_{l,s}$ is created, where l is the ID of a given train line, and s is the ID of a given station. Every resultant set $\mathrm{DT2}_{l,s}$ is composed of villages located in the ROI of station s from line l and straight line distances from the villages to the station. It is worth to notice that particular ROIs may overlap to some extent.

The values of $r_{\min}$ and $r_{\max}$ were selected to be 3 km and 22 km, respectively. Larger values of $r_{\max}$ did not lead to significant changes in the results, i.e. shapes of particular areas in the obtained Voronoi diagrams. However, larger values of $r_{\max}$ significantly increase the number of operations in the subsequent steps. The larger the value of $r_{\max}$ is, the higher the extent to which particular ROIs overlap. Consequently, a given settlement belongs to a bigger number of ROIs. Selected results are shown in Figure 3. Dashed lines connect particular towns with railway stations. This step was also conducted in C++.

Step 3: Computation of distances according to road map

In the third stage (JavaScript), the system added two additional pieces of information to $DT2_{l,s}$ datasets, i.e. the distance along the road and the travel time from each settlement to its station, thus creating respective $DT3_{l,s}$ datasets. It was achieved with the use of Google getDistanceMatrix() service. More than 40,000 combinations were checked for different values of the parameter r. The part of the system developed in JavaScript strongly facilitated the procedure. As the getDistanceMatrix() allows to check only 25 combinations in a single run, the system split the $DT2_{l,s}$ datasets into smaller subsets and provided them to the Google service in the appropriate way (timing).

The resultant $DT3_{l,s}$ datasets provide three values for each settlement-to-station assignment: (i) straight line distance based on geographical coordinates (GEO), (ii) distance measured along the road network (GGLD – GooGLe Distance), (iii) travel time (GGLT – GooGLe Travel).

Step 4: Obtaining the final dataset

Now, for each settlement the system determines the nearest station in regard to three distance criteria: GEO, GGLD and GGLT, combining information from DT1 and $DT3_{l,s}$ to form the final DV4 datasets for each settlement. This computation process may be regarded as filtering with the Min() function.

Step 5: Building Voronoi diagrams

In this step (Octave environment), Voronoi diagrams are finally built for the three above-mentioned criteria. The resultant diagram of the whole Poznan agglomeration for the radius of 50 km is shown in Fig. 4. Fig.5 shows selected areas covered by Voronoi diagrams in regard to the GEO (a), GGLD (b) and GGLT (c) criteria. The results are presented in a coordinate system in which latitude and longitude are coordinate axes in order to provide a legible diagram. Considering a relatively small area (in global sense) of the Poznan agglomeration, it does not introduce noticeable distortions.

IV. Discussion of results

As expected, Voronoi diagram for the GEO criterion unambiguously assigned areas which do not overlap. For the two remaining criteria, particular areas may overlap while being visualized according to geographical coordinates. It is a natural effect, as the roads increase the distances, causing local non-linearities.

Also, Figs. 4 and 5 show the Warta river (thick blue line) which is a natural barrier affecting the shape of Voronoi diagrams. Bridges enabling road traffic are also marked. Exemplary areas on which the river presence results in the assignment of some settlements to different ROIs.

On the basis of such Voronoi maps, it is possible to estimate the boundaries of areas in which autonomous vehicles could be used in public hybrid transport in the future. Factors that should be taken into account are not only the distances to particular railway stations, but also the travel time. An important factor is the total travel time which also includes the train journey to the central city. It is possible that the nearest railway station is located further from the central city, which will additionally increase the total travel time.

One of the possibilities is introducing dynamic shaping zones served by autonomous vehicles based on the travel time, which can be varied during the day, as well as the communication schedule for particular railway lines.

V. Conclusions

The paper presents an idea of improving public hybrid transport system in the future based on autonomous vehicles. Such cars could be appropriate in areas with less traffic, outside of the central city. The goal is to decrease the number of vehicles in already crowded cities by making public transport more flexible.

Another important issue is the appropriate division of the agglomeration area into zones so as to minimize the total distances traveled by autonomous vehicles in a given area. For this purpose, we designed a software system which determines these zones in regard to both road distance and travel time, basing on publicly available data.

The presented results, considered by us as preliminary, are a stage of a larger project. The purpose of the project is to facilitate the planning of public transport in general.

References

[1] I. Azkuna, “Smart cities studies: International study on the situation of ICT, innovation and Knowledge in cities,” *The Committee of Digital and Knowledge-based Cities of UCLG*, Bilbao, 2012.

[2] Spot Mairie, [Online]. Available: http://www.nicecotedazur.org/la-metropole/spot-mairie, (access 2019.05.15)

[3] C. Benevolo, R. P. Dameri, B. D’Auria, (Editors: T. Torre, A.M. Braccini, R. Spinelli), “Smart Mobility in Smart City. Action taxonomy, ICT intensity and public benefits,” *Empowering Organizations: Enabling Platforms and Artefacts, Springer*, ISBN 978-3-319-23784-8, 2016, pp. 13–28.

[4] Bus – Tram – Railway System in Poznan. [Online]. Available: http://bustramwajkolej.pl/

[5] M. Banach, “Use of Intelligent hybrid solutions in sustainable public transportation system for the net of small cities”, *Zarządzanie Publiczne (Public Management)*, No. 3 / 2019

[6] A. Nowak, “Application of Voronoi diagrams in contemporary architecture and town planning” *Challenges of Modern Technology*, Vol. 6, no. 2, pp.30–34, 2015.

[7] M. Abellanas, B. Palop, “Urban Data Visualization with Voronoi Diagrams,” In: Gervasi O., Murgante B., Laganà A., Taniar D., Mun Y., Gavrilova M.L. (eds) Computational Science and Its Applications – ICCSA 2008. ICCSA 2008. Lecture Notes in Computer Science, vol 5072. Springer, Berlin, Heidelberg

[8] F. Bahraminejad, K. Babaki, “Application of Voronoi diagram as an architectural and urban planning design tool,” *Indian Journal of Fundamental and Applied Life Sciences*, ISSN: 2231–6345, 2015 Vol.5 (S1), pp. 1776-1783, [Online]. Available: www.cibtech.org/sp.ed/jls/2015/01/jls.htm

[9] J. Park, S. Ji, H. Jun, “The Application of Voronoi Diagram into the Space Planning for Urban Design,” *7th International Symposium on Architectural Interchanges in Asia* (ISAIA 2008), pp. 524-528.

[10] Z. Xinqi, W. Shuqing, F. Meichen, Z. Lu, Y. Shujia, (2010) “Urban Cluster Layout Based on Voronoi Diagram,” In: Elleithy K. (eds) Advanced Techniques in Computing Sciences and Software Engineering. Springer, Dordrecht

[11] D. R. Mota, M. Takano, Pastor W. Gonzales Taco, “A Method Using GIS Integrated Voronoi Diagrams for Commuter Rail Station Identification: A Case Study from Brasilia (Brazil),” *Procedia - Social and Behavioral Sciences*, Volume 162, 19 December 2014, pp. 477-486.

Communication Papers of the Federated Conference on Computer Science and Information Systems pp. 73–79
DOI: 10.15439/2019F288
ISSN 2300-5963 ACSIS, Vol. 20

Positioning Improving of RSU Devices Used in V2I Communication in Intelligent Transportation System

Marzena Banach[1,2] Katarzyna Kubiak[1,3] and Rafał Długosz[1,4]
[1] Aptiv Services Poland
ul. Podgórki Tynieckie 2, 30-399, Kraków, Poland
[2] Poznan University of Technology, Institute of Architecture and Spatial Planning,
Nieszawska 13C, 61-021 Poznan, Poland
E-mail: marzena.banach@erba.com
[3] Adam Mickiewicz University
Faculty of Mathematics and Computer Science
ul. Uniwersytetu Poznańskiego 4, 61-614 Poznan, Poland
E-mail: kbkubiak45@gmail.com
[4] UTP University of Science and Technology
Faculty of Telecommunication, Computer Science and Electrical Engineering
ul. Kaliskiego 7, 85-796, Bydgoszcz, Poland
E-mail: rafal.dlugosz@aptiv.com, rafal.dlugosz@gmail.com

***Abstract*—In this work we present solutions which aim at enhancement of the localization precision of the road side unit (RSU) devices which will participate in vehicle-to-infrastructure (V2I) communication in future autonomous driving and intelligent transportation systems (ITS). Currently used localization techniques suffer from limited accuracy which is due to various factors, including noise, delays caused by environmental conditions (e.g. temperature variation) and differences in elevation between devices communicating with each other in the road environment. In case of application of the ITS, these factors can be the source of significant discrepancies between real positions of the RSUs and their estimated values provided by the V2I system. The proposed techniques, based on various approximation techniques, as well as linear and nonlinear filters, allow to improve the localization accuracy, reducing the positioning errors by more than 90 %.**

***Keywords*—Intelligent Transportation System, Smart City, V2I communication, Automotive Active Safety**

I. Introduction

An intelligent transportation system (ITS) is a relatively new concept which embraces solutions whose aim is to optimize transport based on modern technologies. They include those in the field of artificial intelligence (AI) as well as information and communication technologies (ICT).

Functionalities considered under the framework of the ITS can be classified into several general classes. One of them are solutions responsible for providing appropriate, up-to date information for passengers of public transport and car drivers. Another group aims at increasing traffic flow in urban areas in order to eliminate or reduce the traffic jams, which can cause a significant reduction in pollution levels in the cities. Finally, one of the main ITS development directions are the vehicles themselves. It is supposed and hoped that the majority of the vehicles in the nearest future will be equipped with advanced driver assistant systems (ADAS). However, future development of the ITS may be understood as the development of autonomous vehicles which will enable traveling without human intervention. The purpose of the development of such vehicles is to improve road safety and ecological aspects.

Taking into consideration the last group of solutions, several essential challenges can be indicated here. As the car moves through the city or suburban/highway areas, the surroundings of the vehicle are dynamically changing. One of the main problems is high complexity of the environment seen by the vehicle. This makes the algorithms responsible for travel safety also complex. The complexity means, for example, a number of objects in the range of the vehicle's sensors, both moving and still objects, their trajectories, obscuring some objects by others, etc. One of the possible solutions in terms of the described problem can be intelligent road infrastructure providing direct support for the vehicle movement, in the form of special devices mounted at fixed points of the road and urban environment.

The support can be provided by the so-called vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication (V2X in short). For example, the V2I system facilitates the operation of traffic sign recognition (TSR) functions. In case of low visibility of traffic signs (TSs), the information passed to the vehicle wirelessly by devices (RSU – road side unit) associated with particular TSs, or groups of them, can significantly improve the performance of such functions. V2I techniques can be used to inform the driver about accidents ahead, bad road conditions, etc. In its most advanced form, a framework of the RSUs participates in building the so-called model (map) of the environment of the vehicle. In the currently

©2019, PTI

used ADAS systems, the map of the environment is created by the vehicles themselves, which use their own on-board sensors (camera, radar, recently LiDAR) and appropriate algorithms. The V2I system can be treated here as an additional sensor which provides additional data, for example from "behind the corner".

One of the main challenges in the described problem is the need of very accurate localization of the RSUs by the passing vehicle. This can be based on a real time localization system (RTLS). Such systems are frequently used in indoor confined areas (buildings) for various purposes. They are usually composed of two types of devices: moving devices (active markers/tags) as well as devices mounted in fixed points of buildings (transponders/anchors). Trajectories of the markers are determined and recorded by the anchors on the basis of multiple distance measurements between particular markers and the framework of the anchors, supported by the trilateration computation techniques [1].

In the indoor systems, the positions of the anchors are usually well known. In case of the urban/road environment, on the other hand, the situation is different. The RSUs can be viewed as counterparts of the anchors, while the markers are the devices associated with the moving vehicles. Contrary to typical indoor applications, it is the device mounted in the vehicle which determines own position in the ITS. Another difference is the lack of fixed framework of the anchors, as the moving vehicle on its path is within the range of different RSUs. Due to these differences, the computation techniques used to determine the relative positions of the vehicles have to be different as well.

We assume that the moving vehicle is able to calculate its own trajectory, within given accuracy, on the basis of its own GPS unit and own on-board sensors (yaw rate and velocity sensors). Due to the noise as well as low precision of the applied sensors, the computed trajectory is also not precise. Constant communication with the RSU devices, which as we assume, know their own positions in the global coordinate system (GCS) increases the precision of the calculated trajectory of the car. The main issue here is how to precisely determine the trajectory in relation to the framework of the RSUs. It is the topic of the presented work.

II. State-of-the art study

In this part we briefly present state-of-the art study in the areas related to the presented topic. As an example of a system which can benefit from the proposed method, we consider the traffic sign recognition (TSR) system which is becoming a standard in the automotive industry, including European New Car Assessment Program (Euro NCAP) [2]. Currently used systems of this type rely only on the on-board sensors (cameras) of a vehicle. Traffic signs (TS) are then identified and recognized using various signal processing and artificial intelligence (AI) methods. The computation scheme is in general similar in each system of this type. Firstly, the TSs are identified in the images taken from the camera, then they are cropped to smaller images and normalized in terms of sizes. On the basis of a series of frames obtained from a single camera or on the basis of stereo-vision cameras, the positions of the TSs in the environment of the vehicle are determined using, among others, the trilateration methods. The cropped and normalized TSs are then provided to an AI algorithm for identification.

A. Towards automotive TSR systems of the 2^{nd} generation

The TSR systems currently offered on the market recognize only selected traffic signs, usually those related to speed limitations, stop signs, etc. In the future, with the development of fully autonomous cars, one can expect that these systems will be able to recognize almost all TSs, similarly to the human driver. This will strongly increase the complexity of the implemented algorithms. The problem which can be frequently observed on the roads is the non-standard appearance of the road signs (damage, coverage, lack of full exposure, suspension at non-standard height, etc.), as shown in Fig. 1. Additionally, taking frequently observed dense arrangement of the TSs over a given area into account, it can lead to various safety issues if the vehicle will only rely on its on-board sensors.

Various efforts which aim at solving the problem of the road sign visibility can be found in the literature. One of the proposed solutions is inclusion of the TSs in the future system supported by the V2I communication. In practice, it means equipping the TSs with the RSU devices capable of transmitting relevant information to passing vehicles. A proposal of such a solution is described, for example, in [3], [4] and [5]. These solutions can be regarded as a next important stage toward the development of active traffic signs.

One of the problems here is how to properly determine the positions of the RSUs associated with their respective TSs in the situation when many TSs are located in close proximity to each other. An exemplary situation of this type is shown in Fig. 2. If the positioning/localization of a TS is not perfect, it can lead to false assignments of the RSUs to the TSs seen by the vehicle and thus to false behavior of the active safety (AS) system of the vehicle.

Precise positioning of the RSU devices will be of great significance for applying them as support in creating a dynamic map of the environment of the moving vehicle.

B. Applications of the RTLSs – desired parameters

The RTLSs are being developed and the localization accuracy is being improved around the world. The majority of these investigations focus on indoor applications [6], [7].

The desired ranging precision (requirements) always depends on the application for which a given system is designed. In some medical applications, even millimeter precision is mandatory. Such a situation takes place in motion capture systems which aim at recording the motion pattern of a disabled person with very high accuracy [8], [9]. Another example is warehouse application for which the precision of 20-50 cm is acceptable. An RTLS of this type is offered by the Ubisense Company [10], with the reported localization errors at the level of 15 cm. This system was designed to operate on relatively large areas. The reported localization errors in

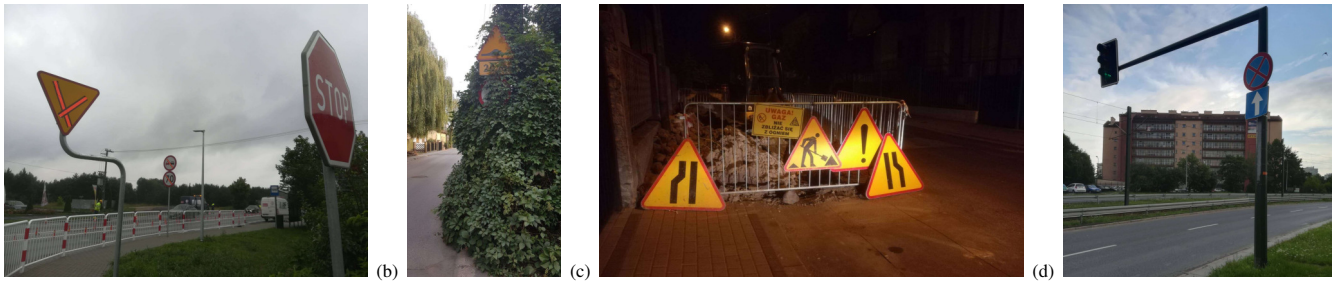


Fig. 1. Examplary real road situations illustrating: (a, b) road signs deviating in appearance from standards or invisible, (c, d) untypical elevation at which a TS can be mounted (own source).

Fig. 2. E

xemplary real road situations illustrating large density of traffic signs in a given area (own source).

indoor conditions are usually below 10 cm [11]. This however, is achievable for relatively small distances of below 10 m. The required precision of the ITS will be different, depending on the target application of the ADAS.

The localization precision is not the only parameter to be considered and optimized. In the above-mentioned indoor human motion capture system the most important problems include improving the ranging precision toward even sub-millimeter range, and miniaturizing the markers to make them comfortable for the examined persons. As these devices are wearable, the energy consumption is also a crucial factor. On the other hand, such systems will be used in the indoor environment, in which temperature is at an almost equal level. Therefore, the robustness to a wide-range temperature variation is rather less important. Normally, in systems of this type, the RTLS tags are affixed to moving objects and tracked by the transponders installed in the fixed points of the environment, with precisely determined (and stored in the memory) positions. The net of transponders may thus create a precise frame of reference for the moving markers. The transponders have access to a power supply (or can be battery operated) and therefore a strong miniaturization is not a critical factor. Additionally, the framework of anchors can take a sufficiently long time to calibrate precisely.

When the RTLS is applied in the ITS, the situation is substantially different from the typical indoor conditions. In the road environment, the network of devices communicating with each other covers a much wider area than in the described indoor systems. Also, the behavior of such systems in cities can differ from the one in suburban areas, mostly due to different device densities in these areas. In the suburban areas, a rather sparse network of the RSU devices is expected, with a small number of devices in the range of the vehicle sensors. This creates some problems in terms of the localization precision, but in some situations it can also simplify the communication scheme (less interference).

C. RTLS-based solutions in Intelligent Transportation Systems

The use of the impulse-radio ultra wideband (IR-UWB) localization technology in an outdoor environment has already been proposed [12], [13], [14], [15]. This technology provides high channel capacity (i.e. the data is transmitted at a high rate) which makes is suitable for use in the V2I communication [12]. Also, propositions to employ the IR-UWB technology in the positioning of the objects on the road have been reported [13].

In sparse and open networks which are expected in the suburban areas, there will be only two devices communicating with each other in the worst case scenario. One of them will be an on-board unit (OBU) of a moving vehicle. The second one – the RSU – will be mounted in a fixed position of the road infrastructure. The trajectory of the moving vehicle can be determined, with a relatively good accuracy, on the basis of the vehicle velocity, acceleration, the yaw rate and other kinematic parameters. A single communication session between the OBU and the RSU will be quick enough to assume that both devices are in still positions during this period, which can be determined using only basis kinematic equations. Let us consider an example: if the distance between both communicating devices is even at the level of 100 m, the velocity is even 100 km/h (22.5 m/s), then the distance traveled

during a single communication session does not exceed 20-25 μm. This value is far below the safety margin and thus can be neglected.

III. Enhancement of the RSU positioning – proposed solutions

In the investigated method, the vehicle measures its distance to the RSU frequently (a time-of-flight (ToF) approach). On the basis of a single measurement session only the distance, r, can be determined, while the azimuth and thus the position of the RSU are not known. As a result, a circle with the radius r is obtained. The RSU is situated somewhere on it. The vehicle is located in the central point of the circle. Theoretically, on the basis of only two measurement sessions (for different car positions) and the trilateration computation, it is possible to determine the position of the RSU in relation to the vehicle.

In practice, various negative factors can interfere with the measurements. Among them are: delays caused by the communicating devices, noise resulting from imperfections of the vehicle on-board sensors, unknown height over the road surface, at which the RSU device is mounted, etc. All these factors result in the measurements being subject to errors. Depending on the source of these errors, they can be either systematic or random.

The problem is illustrated by an exemplary simplified trajectory of the vehicle in Fig. 3. The left diagram shows selected positions of the vehicle, for which distance measurements are performed. The results for different delays introduced by both of the communicating devices are shown. Large black circles illustrate an ideal case, in which delays are known (nominal values). In this case, all obtained circles (marked as "real distances") after a correction by a known factor intersect at a single point, which is the real position (x_R, y_R) of the RSU device.

Theoretically, basing on the trilateration method, the position of the RSU device can be determined from the results of only two measurements, as mentioned above. However, under real conditions, unknown signal delays cause that the resultant circles (marked as "measured distances") intersect in other points (seeming positions). The right diagram in Fig. 3 shows the results for delays deviating from the nominal value towards both positive and negative values. Both types of deviations are shown here for illustration only. In the real situation, only one type will occur. Therefore, in the following figures only single circles are shown for each distance measurement.

The obtained points of intersection (A, B, ..., J) for given values of delay form the area of uncertainty, with the unknown position of the RSU within it, as shown in the right diagram of Fig. 3. The uncertainty at the level of 10 ns can cause the localization error of even ± 1.5 m, assuming the speed of light $v_\mathrm{c} \approx 30$ cm/ns, and two-way communication. Assuming that an error of this magnitude is the case in a suburban area and that the considered system recognizes TSs, then the error could be neglected as the distance between any two TSs is assumed to be significantly greater than the error. In contrast, such error would be too big to be neglected in urban areas or when the RTLS is used to build a dynamic model of the environment.

A. Steps of the proposed enhancement algorithm

The proposed method for estimating the RSU position consists of several stages. In practice, it is an iterative method, where particular steps are performed cyclically and alternately for every new measurement session. The obtained data sets can be kept in memory as a whole or be realized as a delay line (a shift register), as for example in finite impulse response (FIR) filters.

The method described below was implemented in the Octave environment and verified for different values of particular factors and different trajectories of the vehicle. In order to model real conditions in a better way, which is expected during the driving, the measurement values were artificially disturbed by noise with different amplitudes, A, as presented below.

Stage 1: Distance measurements $\rightarrow$ data set with seeming positions of the RSU

For every new, i^th, measurement and the new resultant circle with the radius r_i, new intersection points are computed using the circle determined in the previous iteration.

The objective of this stage is to compute the seeming position of the RSU device (i.e. spatial x and y coordinates). Let us denote the seeming position of the RSU as $\mathrm{SPR}.x, y$. The determined SPRs can be expressed in the global coordinate system (GCS), which simplifies the computations. However, any coordinate system used consistently can be applied as well. The vehicle position can also be expressed in GCS.

It is worth to notice that this method returns a pair of intersection points, so it is necessary to specify which of them is the one reflecting the real position of the RSU. Basing on only two measurements, it is not possible. However, for a larger number of measurements, and thus a larger data set of calculated SRSs, the RSU position can be deduced. If the trajectory of the vehicle is not a perfectly straight line, then the points which are the real equivalents of the RSU position are spatially more focused and form an area which reflects a regular circle (or arc), as shown in Fig. 3. On the other hand, if the vehicle's trajectory would be a straight line, then even larger number of measurements do not resolve the ambiguity problem. Such a situation may happen in urban areas, where straight line movements of the vehicle will be common. Certain amount of curvature in the vehicle's movement is therefore needed. The described ambiguity problem may be solved in other ways, for example through content of messages provided by the V2I system to the vehicle. The vehicles may also intentionally focus only on TSs on one of the road sides.

Stage 2: Seeming positions of the RSU $\rightarrow$ data set with estimated positions of the RSU

The seeming positions of the RSU (SPRs) form an area resembling a circle, with the radius r_SC. What makes the approximation task more difficult is the fact that only a portion of the circle is obtained. How well the points match the circle/arc depends on several factors. A good-quality image of the circle is obtained in the absence of noise and with sufficiently dense measurements. As already described, the

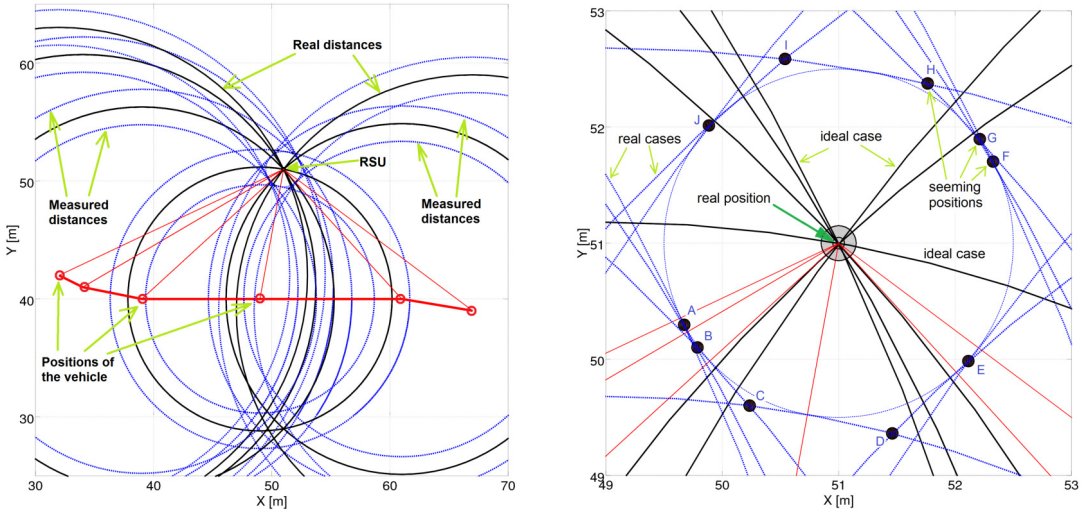


Fig. 3. Determining the RSU position on the basis of multiple measurements - a general idea.

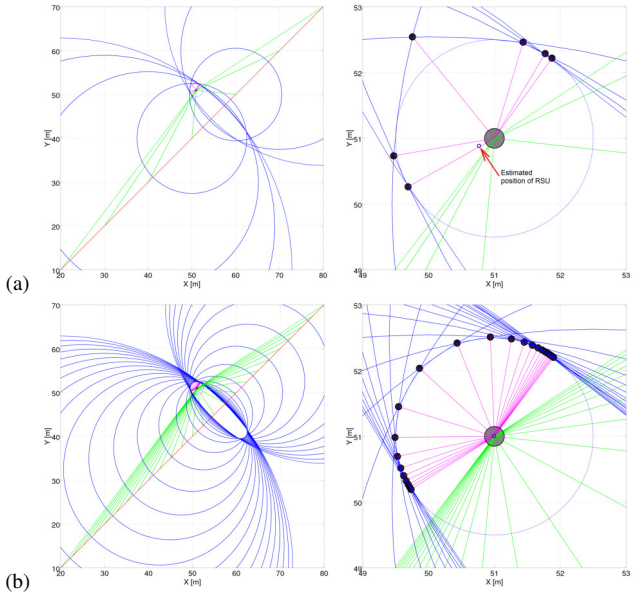


Fig. 4. Selected results for particular measurement frequencies.

objective is to obtain a high density of the SPRs relative to the azimuth angle. If the RSU is far ahead of the vehicle and the vehicle is apporaching it, then the azimuth angle α does not change substantially, and thus the measurements can be less frequent in time. On the other hand, if the vehicle is in a closer proximity to the RSU mounted at the side of the road, then the azimuth varies faster for a given velocity of the vehicle. The investigation results show that $\Delta\alpha$ at the level of 10° is sufficient.

The impact of the selected frequency of the measurement on the accuracy of the computed positions of the SPRs is illustrated in Fig. 4 for constant velocity of the vehicle. In this case the measurements are performed at equal time intervals. In Fig. 4 (a) the measurement density is too small, causing the image of the circle to be distorted. This leads to larger final errors, as marked with the arrow. For a better illustration we set the noise amplitude to zero in this case.

In real situations the noise is non-zero and the SPRs are spread across the edge of the circle. In this situation the parameters of the circle (its radius, as well as the x and y coordinates) have to be determined. This is performed in the second stage of the proposed method. The tests were performed with the noise of different amplitudes. Noise samples were generated according to the uniform distribution with the extreme values between $-A$ and A.

Stage 1 provides the data set with calculated SPRs. Stage 2 introduces an approach which relies on estimating the circle parameters basing on three selected points, which are assumed to be located on this circle. In case of a high angular density of the distance measurements, particular SPRs can be located very close to each other. To minimize the impact of the noise on the estimation results, the points are selected so that they are not too close to each other. For the purposes of the presented method, existing methods of this type were adapted and some modifications were introduced.

Stage 3: Filtering over the data set of the estimated positions of the RSU

Stage 2 provides a new dataset with computed intermediate seeming circles, representing estimates $(x_{\mathrm{E}}, y_{\mathrm{E}})$ of the real position of the RSU $(x_{\mathrm{R}}, y_{\mathrm{R}})$. The size of this set can increase with each new distance measurement or be constant as in the delay line used in filters. Depending on the amplitude of noise, the obtained intermediate circles can significantly differ from each other, in terms of both the position and the radius. The objective of this stage is to find an appropriately averaged circle, with a supposed real position of the RSU.

In this (final) stage various operations can be performed including: sorting, filtering, truncating, etc. Multiple tests (10,000+) allowed us to determine a method and its settings

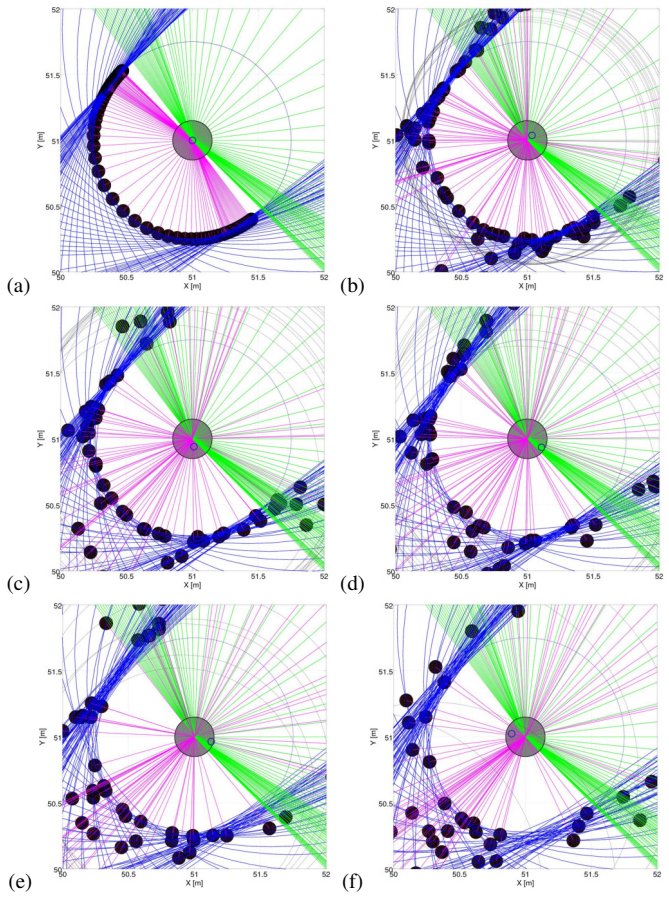


Fig. 5. Estimated positions of the RSU for time delay of 5 ns (1.5 m) and different values of the noise (amplitude A) applied to the measured distance, for : (a) $A = 0$ [mm], (b) $A = 150$ [mm], (c) $A = 300$ [mm], (d) $A = 450$ [mm], (e) $A = 600$ [mm], (f) $A = 750$ [mm].

which provide a relatively good approximation of the positions of the RSU devices, including the case of noise with relatively large amplitude.

The dataset which results from Stage 2 can contain biased data (i.e. circles with very large radius values). Therefore, we initially remove all circles whose radii are greater than fixed radius threshold value (for example, 3 meters) in order to ensure the reliability of the data. Also, the extreme case when all circles are larger than the fixed threshold should be not be analyzed. The threshold value can be adjusted on the basis of statistical analysis of the obtained results.

After the initial data verification we sort the remaining dataset separately over the x and the y coordinates of the middle points of the circles. We observe that usually a large portion of the dataset with the circles is located in the proximity of the real position of the RSU. After the sorting operation, these circles are located in the middle of the dataset, for both the x and y coordinates, while those that have totally false values (in terms of x and y coordinates) are located on both peripheral areas of the set. In one of the tested scenarios, a smoothing low pass FIR filtering over the sorted dataset was applied. In these tests, filters with equal coefficients of the order N = 1, 2 or 3 were used. This operation slightly improved the results. However, to come to a more general conclusion and to avoid false positives, further investigations are required.

The sorted and optionally filtered datasets are truncated at both sides, taking into account some additional parameters. One of them is the spread (threshold) across the median value of the remaining dataset. The threshold can be set to be a constant parameter or be determined, for example, on the basis of the variance across the median value.

Finally, the x_{E} and y_{E} (E – estimated) coordinates of the RSU are found as the median values of the output dataset. The calculated position is still only an approximate of the real position. However, the obtained results, shown in Fig. 5, are substantially better than those expressed only by the SPRs, computed in Stage 1. Comprehensive investigations performed for different values of particular parameters (trajectories of the vehicle, noise levels, delay times, etc.) show that the determined positions can be improved by more than 90 %. This value was determined by comparing two quantities for each performed test. One of them is the radius of the seeming circle, r_{SC}, resulting from the assumed signal delays, shown in Fig. 3 (right). The second one is the distance between the real position (x_{R}, y_{R}) of the RSU device and the estimated position (x_{E}, y_{E}) determined by the applied method. For a single test, the precision, P, is expressed as follows:

$$P = \frac{r_{\mathrm{SC}} - \sqrt{(x_{\mathrm{R}} - x_{\mathrm{E}})^2 + (y_{\mathrm{R}} - y_{\mathrm{E}})^2}}{r_{\mathrm{SC}}} \qquad (1)$$

The obtained precision of 90 % is based on multiple tests and is the value for the worst-case scenario.

A good practice is to try to reduce the amplitude of noise directly at the outputs of the on-board sensors of the vehicle responsible for calculation of its trajectory, for example, of the yaw rate sensor. This substantially improves the results and simplifies the computations in Stage 3 of the proposed procedure. We have already dealt with this problem during the implementation of a company automotive project, in which it was necessary to precisely filter the yaw rate signal to improve the behavior of realized active safety functions [16].

IV. Conclusions

In this work we presented solutions which can be of importance for the implementation of novel automotive ADAS functions and their common usage in smart cities and in the ITS in the nearest future. High quality standards for such systems is expected, as the goal is to ensure full vehicle autonomy relatively soon. The obtained results are therefore important as the safety of the Intelligent Transportation System users is concerned. It is frequently stressed that this system will need support from the urban and the road infrastructures. One of the main challenges involved with this idea is the precise determination of the vehicle position with respect to the urban infrastructure, other road users and other objects.

A prototype of the proposed method was implemented in the Octave/Matlab environment and tested for different values of particular parameters (trajectories, noise, measurement frequencies, etc.). The model was investigated for different values

of the noise and time delays – the sources of the systematic and random errors – and it was shown that the estimation of the real position of the RSU can be improved by more than 90 %. For test purposes, we selected relatively large values of particular distorting factors, which are rather not expected in the real environment – the radius of the area of uncertainty greater than 1.5 m, accompanied by the noise with the amplitude at the level of 50 % of the selected radius. After the calculations, the obtained deviations from real position do not exceed 2 – 10 cm., which can be regarded as a satisfactory result.

Larger values of distorting factors were chosen for one more reason. Real road tests are needed to verify the reliability of the proposed method. Presently, it is not possible to carry out comprehensive tests of this type due to the lack of functioning V2I networks. That is why larger values of distorting factors were assumed to take the worst case scenario into consideration.

However, all calculations were performed assuming equal heights of the RSU and the vehicle. The problem of unequal heights would require additional calculations in three dimensions. Another solution would be equipping the devices with the information regarding their heights, which would result in less complex computations.

V. Acknowledgement

The presented work is a part of the project: "Development of innovative technologies in the field of active safety, which will be used in advanced driver assistance systems (ADAS) and autonomous driving systems intended for mass production", under the framework of "Quick path" program, No. POIR.01.01.01-00-1398/15 (Poland).

References

[1] O. Hassan, I. Adly, K.A. Shehata, "Vehicle Localization System based on IR-UWB for V2I Applications", *8th International Conference on Computer Engineering & Systems* (ICCES), Nov. 2013.

[10] Lianlin Zhao, Eric T. Psota, Lance C. Pérez, "A Comparison Between UWB and TDOA Systems for Smart Space Localization", *IEEE International Conference on Electro/Information Technology* (EIT), 2014, pp.179-193.

[2] Euro New Car Assessment Program (NCAP) – Safety Assist: https://www.euroncap.com/en/vehicle-safety/the-ratings-explained/safety-assist/

[3] Lin Mingyao, "Transmission system for wireless electronic traffic sign and position", Patent CN1841411 (A) – 2006-10-04.

[4] Hu Jianguo, "Traffic sign wireless broadcasting system", Patent CN203070537 (U) – 2013-07-17.

[5] Kuo Hsiao-Chen, "Traffic sign wireless broadcast device", Patent TW201104632 (A) - 2011-02-01

[6] Marzieh Dashti, Mir Ghoraishi, *et al.*, "High-Precision Time-of-Arrival Estimation for UWB Localizers in Indoor Multipath Channels", *Novel Applications of the UWB Technologies*, edited by Boris Lembrikov, ISBN 978-953-307-324-8, Aug., 2011.

[7] Ł. Zwirełło, *Realization Limits of Impulse-Radio UWB Indoor Localization Systems*, Dissertation, Karlsruher Institut für Technologie (KIT), Fakultät für Elektrotechnik und Informationstechnik, 2013.

[8] P. Merriaux, Y. Dupuis, R. Boutteau, P. Vasseur, X. Savatier, "A study of Vicon system positioning performance", *Sensors*, 17(7), 1591, doi: 10.3390/s17071591, 2017.

[9] R. Długosz, J. Pauk, P.A. Farine "New Trends in Motion Capture Systems for Human Gait Analysis", *Machine Graphics and Vision*, vol. 20, no. 3, 2011, pp. 319-331.

[11] P. Karbownik, G. Krukar, A. Shaporova, *et al.*, "Evaluation of Indoor Real Time Localization Systems on the UWB Based System Case", *2015 International Conference on Indoor Positioning and Indoor Navigation* (IPIN), Banff, Alberta, Canada, Oct. 2015.

[12] S. Kabil, B. Ait Essaid, A. Ait Ouahman, *et al.*, "Analysis of UWB-OFDM system for Vehicle to Infrastructure communication", 4^{th} *International Conference on Logistics* (LOGISTIQUA), June 2011, pp. 319-331.

[13] D. Dardari, N. Decarli, *et al.*, "High-Accuracy Tracking Using Ultrawideband Signals for Enhanced Safety of Cyclists" *Mobile Information Systems*, Vol. 2017, Article ID 8149348, 2017.

[14] S. Kabil, R. Elassali, F. Elbahhar, *et al.*, "Orthogonal Frequency-Division Multiplexing Ultra Wide Band System in Real Environment for Vehicle to Infrastructure Application", *Journal of Computer Science*, Vol. 9, Iss. 10, DOI : 10.3844/jcssp.2013.1305.1317, pp.1305-1317.

[15] J. Wang, Y. Gao, Z. Li, X. Meng, C. M. Hancock, "A Tightly-Coupled GPS/INS/UWB Cooperative Positioning Sensors System Supported by V2I Communication", *Sensors*, 16(7), 944; doi:10.3390/s16070944, 2016.

[16] R. Dlugosz, M. Szulc, M. Kolasa, *et al.*, "Design and Optimization of Hardware-Efficient Filters for Active Safety Algorithms," *SAE International Journal Passengers Cars – Electronic and Electrical Systems*, 8(1):2015, doi:10.4271/2015-01-0152, 2015.

Communication Papers of the Federated Conference on Computer Science and Information Systems pp. 81–88
DOI: 10.15439/2019F329
ISSN 2300-5963 ACSIS, Vol. 20

Design of a Distributed HIDS for IoT Backbone Components

Guilherme de O. Kfouri*, Daniel G. V. Gonçalves†, Bruno V. Dutra‡, João F. de Alencastro§, Francisco L. de Caldas Filho¶, Lucas M. C. e Martins‖, Bruno J. G. Praciano**, Robson de O. Albuquerque††and Rafael T. de Sousa Jr.‡‡
National Science and Technology Institute on Cyber Security, Electrical Engineering Department, University of Brasília (UnB), P.O. Box 4466, Brasília–DF, Brazil, CEP 70910-900
Email: kfouri@ieee.org*, { daniel.goncalves†, bruno.dutra‡, joao.alencastro§, francisco.lopes¶, lucas.martins‖, bruno.praciano**, robson††, rafael.desousa‡‡} @redes.unb.br

***Abstract*—Nowadays DDoS attacks using devices from IoT networks are frequent and extensive. Given that IoT network instances are distributed and deployed on the conventional Internet structure, DDoS countermeasures in IoT need to be fully distributed and coordinated all over the components that form each IoT instance. This paper presents a host-based intrusion detection system (HIDS) that was designed and prototyped to protect the components of IoT network backbones comprising conventional switches and routers, not IoT devices. In our design, a set of the proposed HIDS executes conventional security verification, like default username and password, known attacks signatures, usage of resources, processes, ports and open connections, while also interacting with a Controller of the HIDS set to allow the coordination of intrusion detection actions relative to DDoS attacks distributed all over the IoT instance. The designed distributed HIDS is evaluated in a controlled environment that, although being a local and isolated network, realistically represents IoT network instances.**

***Keywords*—Internet of Things (IoT), IoT Security, Distributed Denial of Service (DDoS), Host-based Intrusion Detection System (HIDS), Distributed HIDS, Mirai botnet.**

I. Introduction

Nowadays a significant increase has been observed in both the total number and duration of distributed denial-of-service (DDoS) attacks [1], one of the most known cases being the Mirai botnet [2], which on October 21, 2016, performed a massive attack that left much of the US East Coast without internet access. This attack broke the record of generated traffic in that year reaching 1.2 Tbps using a worldwide network of security cameras and leaving websites such as Netflix, Twitter, Amazon and PayPal unavailable. This record was latter surpassed by a 1.35 Tbps attack against GitHub [3].

This kind of malware is particularly noted for taking advantage of IoT devices which are plagued by basic security vulnerabilities, such as default user identifiers and passwords. This allows the malware to establish access and then to be installed, thus turning the device into a bot victim or zombie, i.e., a malware-infected device that is somehow controlled by an attacker [4]. Upon reaching the desired amount of bots, the botnet controller commands the attack that is performed from potentially all compromised bots. Then, the DDoS attack consists of bots sending requests and traffic to a particular server to exhaust its computational resources and to exhaust the limit of connections supported by the server, disabling it to function properly.

These attacks, including the Mirai botnet and its variants, as well as other different botnets, call for greater security in IoT devices and IoT network components, given the risk of exposing the Internet infrastructure to increasingly larger DDoS attacks [4]. Popular cybersecurity reports, such as [5], discuss the reasons for such security events highlighting that organizations and users are implementing low-cost IoT devices as quickly as possible with little or no security concern. There is also the possibility of new vulnerabilities being discovered even if the known ones are repaired. In this situation, permanent and evolving security measures must be integrated into the components of IoT network instances.

Given the need for this type of security measure for IoT networks, and considering that current IoT network infrastructures mostly use conventional routers and switches, this paper proposes a distributed host-based intrusion detection system (HIDS) for IoT backbone components, including conventional switches and routers. IDS are security tools that are becoming increasingly necessary as firewalls and other security measures are not sufficient to guarantee the integrity of the network [6]. In our design, a set of local HIDS is distributed into the IoT backbone and these local HIDS interact with a set Controller that coordinates the distributed HIDS.

The proposed design also considers that a basic action to identify security vulnerabilities is the proactive monitoring of network elements. Several protocols can be used for this purpose, such as the traditional monitoring via the Simple Network Management Protocol (SNMP) that allows a network manager system to perform periodic requests from local Agents to network devices. This protocol is leveraged in this paper proposal. Another protocol used to support the detection of security vulnerabilities is Syslog. It is used to convey event notification messages [7] that usually come from logs generated by the operational systems, firewalls or other network elements and contain information about events in each host. The logs can be accessed locally or remotely or can be exported via Syslog to a centralized station where they can be analyzed and stored for longer periods.

Our design uses proactive monitoring allied to intrusion

©2019, PTI

detection to allow the identification of malicious attacks and reduce significantly the possibility of invasions. The proposed HIDS is designed to be integrated with the IoT middleware and its supporting components as a fully distributed security service, given the need to have coordinated protection covering the whole set of components present in each IoT instance, as discussed in [8].

Besides this introduction, this paper is organized as follows: Section II presents a brief review of literature about IDS and Section III presents related works on security issues in IoT backbones. Section IV is devoted to the proposed Distributed HIDS for IoT Backbone Components and Section V describes respective validation tests and their results. Finally, Section VI presents general conclusions and suggestions for future work.

II. Related Works

As this paper focuses on security issues in IoT network backbones, this section summarizes literature views of some concepts that are useful for understating the paper content.

A. IDS

Authors in [9] define intrusion detection systems as a combination of software and/or hardware components that monitors computer systems and raises an alarm when an intrusion happens. Other references, like [10], include in their definition the detection of policy violation and the logging of events.

Intrusion detection systems are a very common tool in the defense against multiple kinds of threats, being widely utilized because of their benefits of flexibility, effectiveness, and interoperability, although these systems cannot be used as the sole and complete solution for security problems, as discussed in [6].

Existing IDS proposals use different structures and processes, regarding the location of the IDS, detection methods, responses, timing, architecture and criteria for monitoring targets. Still referring to [6], a Network-based IDS (NIDS) is the most common form of intrusion detection system whose basic idea is to monitor and analyze packets mirrored from a switch, firewall or any other network active device, and the alert about possible malicious activity. Alternatively, a Host-based IDS (HIDS) analyzes data internal to a computer system, such as audit trails, system logs, and critical system files.

B. HIDS

According to [11], an HIDS monitors and collects the state of hosts or server systems that are running in public services containing sensitive information, and the related suspicious activities. Reference [6] shows that HIDSs have a high precision rate on determining users and processes which are involved in an attack. But, paper [10] points out a potential disadvantage, since HIDS rules are predefined based on the current operational systems architecture and its behavior, thus future upgrades and drastic changes could cause problems.

C. Remote IDS

A remote IDS works similarly to a normal IDS, apart from the information being transmitted over the network to a resourceful server. Paper [12] proposes a system built using distributed intelligent data mining agents. Those agents would exchange the gathered data, logs and activity with each other and detect malicious activity at different levels.

III. Security Issues in IoT Backbones

Also in this section discusses other works about security systems for network infrastructure components, e.g., switches and routers, and also points out the difference between their approach and our proposal.

The security of network gear has been an interesting research and development issue for academy and industry given the permanent concurrency regarding ever evolving security requirements for network devices. As a result, for instance, devices may come with the capability of port security [13], which is a router software configuration that enables the filtering of packets via a white list of previously configured middle access control (MAC) addresses that are allowed to communicate with a given interface. In this case, when a packet from a not-allowed MAC source arrives in an interface with port security configured, the router software will apply a pre-configured action like port shutdown. Port security, however, is not the only security configuration for network devices, Direct Host Configuration Protocol (DHCP) snooping [14] is another method for implementing security in the network infrastructure that prevents a fake DHCP server from distributing IPs on a target network. This type of protection is of great importance in the network because the false DHCP server can distribute IPs inside a valid network range which will eventually generate duplicity of addresses, and thus unavailability of services. These solutions are efficient ways to protect the network devices against a small set of attacks. However, for other attack types, such security settings are useless, and there are still operating devices that do not have such security capabilities. Our solution differs from the discussed ones since its purpose is to complement these security policies by employing the analysis of collected metrics from the network devices operating system, regarding known security flaws for specific devices, malformed passwords, leaked passwords, SNMP communities etc, offering the functionalities that are discussed in Section IV.

Some authors have addressed secure network device leveraging an IDS in different ways. For instance, in [15], the authors propose an IDS called JiNao that analyzes OSPF traffic arriving in a network router to detect deviations from the protocol expected behavior. This detection system is divided into four main entities:

1) Rule based prevention: Module that implements a set of rules/policies used to filter packets related to a previously known violation. The application of these rules serves to avoid attackers to cause low response time from the IDS due to unnecessary processing;

2) Detection module: this module is responsible for inspecting OSPF packets to find deviations in the protocol behavior, which is performed according to two main methods: protocol and statistical analysis. The protocol analysis is implemented with an OSPF state machine that describes the normal OSPF operation and allows to detect whether an OSPF entity goes to a state that doesn't match the expected one. In this case, an alert is generated. The statistical analysis is used to define a behavioral pattern for the OSPF protocol so that, if a packet deviates from the pattern, an attack is identified. These methods are complementary because they address different situations;
3) Decision module: this module is the responsible for decision making in the JiNao IDS, a process that takes the output from protocol and statistical analysis to determine if an attack is happening and then, if misbehavior is identified, an alert is generated destined to responsible entities;
4) JiNao MIB: the IDS maintains an SNMP Management Information Base (MIB) to provide network management services with the information about the security analysis regarding the OSPF protocol. This MIB is used to feed information generated by JiNao into a network management system like Zabbix.

This organization of JiNao provides a security system for the network infrastructure that acts directly in the OSPF routing protocol, verifying if the protocol execution is as expected. This implementation however only focuses on the routing protocol, staying limited to the network layer operation, while our proposal acts directly in the network infrastructure system to detect vulnerabilities.

Also regarding network infrastructure security, the paper [16] proposes a NIDS for active routers using the data mining algorithm random forest for intrusion detection. First, this NIDS captures packets from the network and these packets are submitted to the random forest algorithm aiming at the creation of a dataset for the network existing services. This dataset is then used as the default network behavior to be used for the detection of intrusions which are characterized as anomalous behavior when compared to the created dataset. This cited paper argues that, due to the level of complexity of the misbehavior detection algorithm, the traffic analysis can not be made as the packets arrive, so the detection happens sometime after the packets arrive in the security system. In this same paper a three-layer architecture is proposed to enhance the security of the network routers, comprising a protocol decoder, a misuse detector and an anomaly detector. This anomaly detector is based on network behavior analysis using the previously generated dataset. The misuse detection proceeds with serial packet payload analysis involving each of the network routers, based on the fact that the first-hop router produces a message digest of the packet payload using the Message-Digest algorithm (MD5). Thus this digest is analyzed in each network router to see if the packets have been tempered with. Although this approach detects misbehavior by comparison with the standard network services pattern created using random forests, it is, however, unable to detect problems occurring in the router host as our proposal which acts directly in the network device to detect its security flaws.

Regarding host-based IDS proposals, the work [17] introduces the idea of a Distributed Intrusion Detection System (DIDS) which gathers information from remote hosts on a local area network (LAN), being directly related to the work being done in this paper. In the cited paper, a central role is given to the director entity, which represents the system where the information converges to. Similarly, in our proposal a centralized station, the Controller, collects and processes information gathered from the hosts of the network infrastructure, but while the LAN communications in [17] uses the CMIP protocol, which is an application layer protocol that lacks internal security, our proposed architecture leverages the SSH protocol and its safeness since it assures message authenticity by distributing the public key of the centralized station to the cooperating hosts in the proposed configuration process.

A remote IDS is proposed in [18] by means of web service that provides data analysis to hosts that do not have a local IDS. In this configuration, the hosts and sensors collect preprocess data that is sent to the centralized data analysis web service, which has a role similar to the Controller in our proposal. It is worth to point out that our centralized station is able to processing the data from a set of hosts and uses a neural network to analyze and perform machine learning so it allows more precise analysis and countermeasures as new agents use the IDS.

IV. Distributed HIDS for IoT Backbone Components

This work proposes the implementation of a Distributed HIDS for IoT Backbone components. In view of the concern with the security of IoT networks, the HIDS proposed must be located in a very usual infrastructure nowadays, an IoT network that has conventional routers and/or switches to connect the specific IoT equipment. So this paper focuses on modules for these routers and switches, not IoT devices, considering that complementary protections for the IoT specific devices must be provided with countermeasures specific to these devices such as proposed in the previous and correlated paper [19]. This cited paper describes a local signature-based HIDS that runs in IoT smart devices. Once a Smart device is inserted in the network, it must download the HIDS application. Then, each device updates its set of local rules with the remote rules that are constantly revised, maintaining all the final devices up-to-date with the security requirements. Thus, this paper proposal must be considered as being part of a comprehensive security solution for IoT in which different IDS configurations are established among different layers of the IoT infrastructure.

The architecture of the HIDS proposed in the present paper is shown in Fig 1 with the backbone components, i.e., conventional switches and routers in an IoT network instance,

and our proposed Controller present and available by means if the Internet infrastructure. The Controller maintains databases for the data coming from the backbone components, as well as for the rules that will be used by the distributed detection set of HIDS.

The HIDS will remotely perform a set of periodical security verifications in each of the network managed devices and generate a report on the security vulnerabilities that are found in routers or switches.

To perform this role remotely, a Controller uses the database containing the necessary information collected from the backbone components that must be checked and the other database containing the rules for the verification to be executed. The database for backbone components data stores the information according to the JSON format, as shown in Fig. 2, so it can be easily used as a Python dictionary. Passwords are saved as hashes in this database.

A. HIDS parameters

The HIDS development was intended for the network administrator usage, as it needs confidential information from the backbone components. The information needed is:

- device type
- IP address
- username
- password
- secret

The Controller allows the administrator to list the devices that are registered in the database and then the administrator is able to insert or remove a device. The Controller will connect to each listed device through an SSH channel and run authenticated verifications. The administrator also has the option of not revealing the passwords of the backbone devices, then the Controller will execute some unauthenticated verifications. In this case. as the Controller cannot establish the connection to the device, it will try a brute force access to discover the username and password. If it is successful then all the verifications will be executed, otherwise just CPU usage and port scan verifications will be done.

In the end, a report is generated containing the detected security vulnerabilities highlighted in red and good security practices highlighted in green. Each device will have its specific report so that the administrator is able to take actions to correct security vulnerabilities considering all of those reports.

B. Security Verifications

The Controller performs a series of analysis on the remote backbone component to verify the integrity in the device's system. These verifications are presented bellow and classified according to the related risk.

1) Default username and password: One of the most common vulnerabilities is the use of default or common user identifier and password in devices. These devices are extremely vulnerable to malware that uses brute force dictionary attacks to gain access to a device. For instance, the malware Mirai malware uses a simple dictionary for the default username and password [2].

As mentioned above, as our proposed verifications will also be done when the administrator does not reveal the device password, in these case the Controller will use one brute force dictionary and try to establish the connection to the device. If the connection is successful it indicates that the device is vulnerable and has a default or common username and password.

In the case the administrator provides the device password registered in the Controller database, the connection will be established using the given password. Therefore, for the verification, instead of trying to establish a connection to the device using a brute force dictionary attack, the Controller will access all usernames and passwords in the device. For instance, ss the passwords are encrypted in usually 2 types in Cisco switches, type 5 and type 7, the Controller will try to convert the hash password to plain text, and then compare it with a brute force dictionary. If the password is successfully converted to plain text and is in the dictionary, it indicates a default or common username and password vulnerability.

It is important to emphasize that the data acquired in the brute force attempt is only used by the local Controller to find security vulnerabilities and remains in it. In no way, the information gathered will be used against the owner of the device, as it is expected that the Controller computer is physically and logically protected.

Based on the fact that a potential attacker can use the default username and password vulnerability to fully compromise the confidentiality, integrity, or availability of a target system, particularly in IoT instances, it is classified as a high-risk vulnerability.

2) Leaked password: Some attacks on companies that have a database of user information often lead to a leak of passwords, credit card numbers among other confidential information. Given this type of leak, a device password check is performed on the "haveibeenpwned.com" site.

The site checks if a password has already been compromised in some previous type of data leak. To not expose the password by sending it to the site, a hash technique is used. In this technique the user sends a request to the site with the first 10 characters of the password after it has been encrypted, then the site returns a list with the compromised passwords corresponding to the 10 hash characters sent. Then a local comparison is made to know if the user password is within the list, if so it implies that the password has been compromised by some leak.

Again it is important to highlight that the information will be used only locally and made available only for the administrator of the network to take actions.

As described for the precedent vulnerability, the flaw now described also creates loopholes that can completely compromise the integrity of the system. Therefore, it is also considered a high-risk vulnerability.

3) Known vulnerabilities: The National Vulnerability Database (NVD) provides to the community a large list of

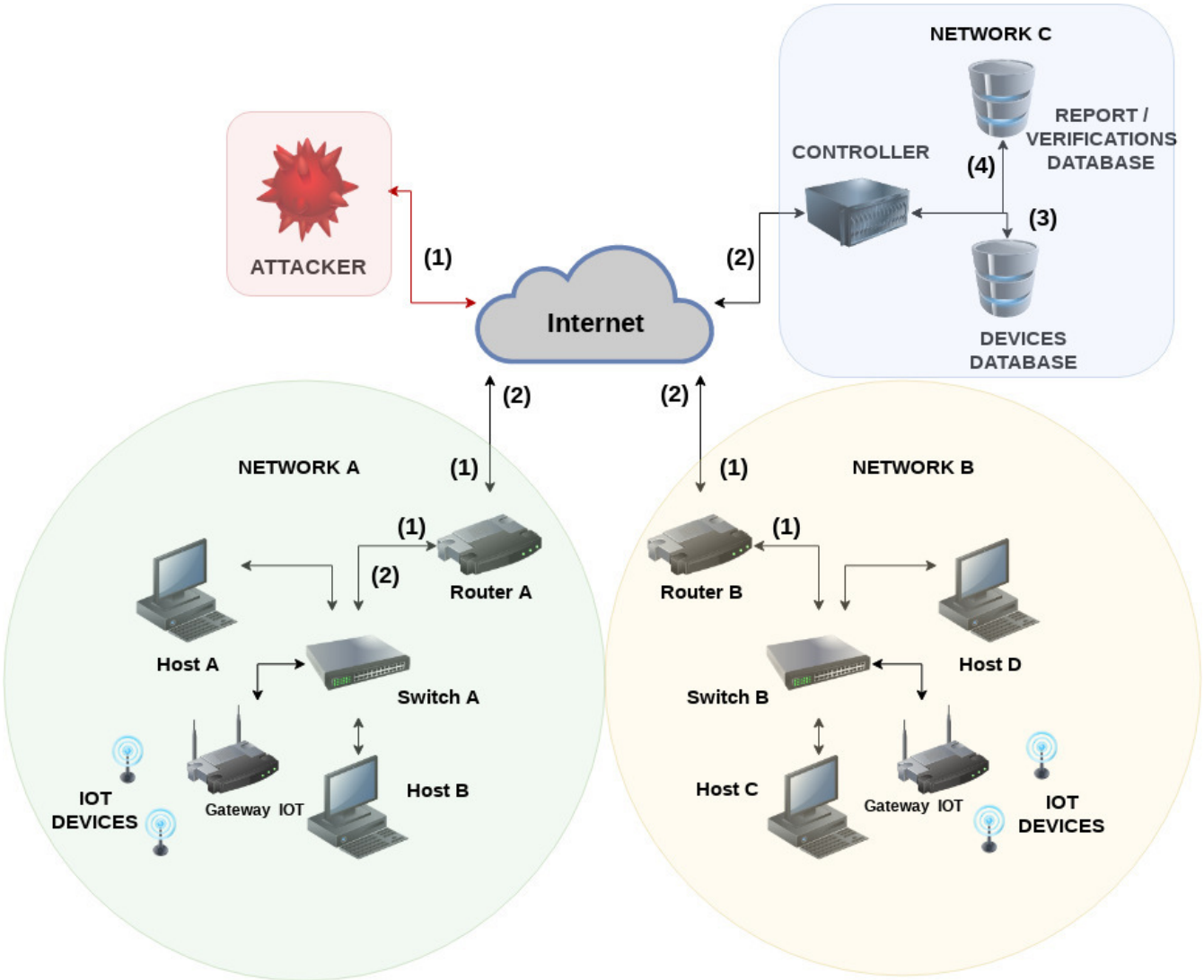


Fig. 1. The Logical architecture of our proposal. 1 represents the the attacker entry point. 2 represents the connection between the Controller and the network devices. 3 and 4 represent the data flow between the Controller and its databases

```
{
    "id"            :   1,
    "device_type"   :   "cisco_switch",
    "ip"            :   "10.0.2.2"
    "username"      :   "switch¨,
    "password"      :   "$1$bM64$SHZqwrzUqBN/0zKjeHiVB1",
    "secret"        :   "$1$J5fz$L0lO0ZHrWlsAd72g/qP8h1"
}
```

Fig. 2. JSON of a backbone component

security vulnerabilities, classified by the vendor, the model and the version of systems. For each vulnerability, a Common Vulnerabilities and Exposures (CVE) is created. CVEs are used to identify and catalog vulnerabilities in software or firmware into a free dictionary for organizations to improve their security.

The Controller, knowing the vendor, the model and the version of the device, as gathered by the SNMP protocol, is able to request all the CVEs for each specific device being verified. Then it gets a list of the CVEs containing the CVE-id, the publication date and a description of the vulnerability. This information with data obtained in the verifications allows to determine if there is a vulnerability identified by a CVE for each device. This information will be given to the administrator to take action.

As these vulnerabilities are extremely variable, their severity depends on the degree of risk they represent in a specific routing configuration and can not be previously classified.

4) Resources used by routing devices: Resources used by a router or a switch may indicate a problem with the device or some possibly malicious action that is trying to overload it.

Therefore, in our proposal a check is provided for the amount of CPU and memory being used by the devices.

Using the SNMP protocol, the percentage of CPU usage is checked and, if it is above 80 percent, this indicates that something is compromising the performance of the device. Then, a scan of the running processes is performed to list the processes that are consuming more CPU and possibly are related to the cause of the problem.

The amount of memory is also checked, including he total amount, the amount used and the amount of free memory. If the percentage of usage is higher than 80 percent of the total amount, the report will flag it as a potential vulnerability.

These vulnerabilities do not always represent an imminent risk to the integrity of the system as a whole and may sometimes only represent the normal functioning of the system. Therefore, it was considered a low-risk vulnerability.

5) Port scan: The device ports are scanned to check for possible security vulnerabilities in open ports. This information will be used in parallel with information already known from the model and specific version of the device.

Often port scanners are used in attack planning phases and can lead to serious security incidents. Having open ports is considered a medium risk vulnerability.

6) Current configuration: Once the Controller has established a connection to the backbone component, it will analyze the current configuration on the device. It will look specifically for configurations that can lead to misbehavior on the device.

The Controller starts by checking if the SNMP community string is common and whether it gives write permission to the MIB, which would allow an attacker to make changes in the device, what is considered an indeed important vulnerability. The Controller then looks if the HTTP protocol is enabled in place of HTTPS, as HTTPS is more secure than HTTP. Then it is checked if the IP CEF is enabled, as if not it can cause slowness in the device.

Finally, the Controler analyses if the telnet protocol is enabled, as using this information with the known vulnerabilities for the specific model of the device it is possible to classify it as a problem or not in the device.

Having the write permission to a SNMP protocol grants power enough to change specific configurations and to even turn off a device.Therefore, it was considered as a high-risk vulnerability.

Table I summarizes the risk rating of each vulnerability classified by the proposed IDS.

V. Tests and Results

To validate our Distributed HIDS for IoT Backbone Components we developed a corresponding prototype and performed a battery of tests in a controlled environment, i.e., a local and isolated network without compromising public or other private backbone components.

We executed the developed software to evaluate 3 backbone components. Fig. 3 shows the architecture used in the tests.

TABLE I
VULNERABILITY RISK RATING

Vulnerability	Risk
Brute force	high
Common password	high
Leaked password	high
Common secret	high
Memory utilization	low
CPU utilization over the last minute, CPU utilization over the last 5 seconds, CPU utilization over the last 5 minutes	low
Community string in SNMP protocol	high
IP CEF	low
HTTP is enabled without HTTPS	medium
Telnet enabled	medium

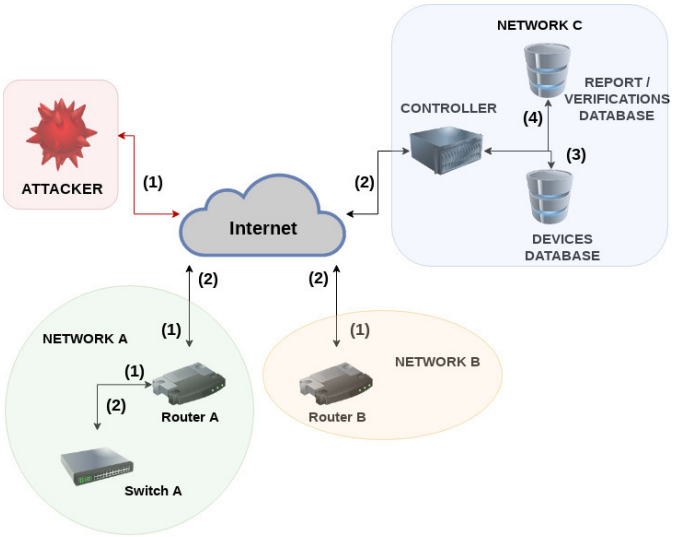


Fig. 3. Simplified architecture of the tests executed.

TABLE II
EXPECTED FORMAT OF THE OUTPUT IN THE REPORT

Vulnerability	Format
Brute force	If it was used, if it was successful
Common password	Shows the common passwords found separated by comma
Leaked password	The password: amount of occurrences
Common secret	Default secret or no secret
Memory utilization	Memory utilization in percentage
CPU utilization over the last minute, CPU utilization over the last 5 seconds, CPU utilization over the last 5 minutes	CPU utilization separated by comma
Community string in SNMP protocol	The common strings followed by the configuration separated by comma
IP CEF	Configured or not
HTTP is enabled without HTTPS	Yes or no
Telnet	Open or not
Port scan	Open ports separated by comma
CVE Vulnerabilities	CVE-id, date, description

The first backbone component, called 'Device 1' is a Cisco Router where almost all the security vulnerabilities are present, except the exhaustion of resources which we could not simulate. The administrator has given the SSH device password and username to the IDS, so that the brute force technique is not utilized.

The second backbone component analyzed by our HIDS was the Cisco Switch named 'Device 2'. In this case, the administrator do not give the SSH username and password to the IDS, so it utilizes the brute force dictionary trying to guess the credentials. As it is successful messages about default username and password are printed in the report, as shown in Table III. Now, having the credentials of the device, the HIDS continue to execute the other verifications.

The third backbone component is a Cisco Router named 'Device 3' and considered by the authors a router with configurations close to those expected in the majority of real environments, which means that the vulnerabilities found in it are also the ones expected by the authors to be the most present when implementing the HIDS in a realistic situation.

Having the information needed from the 3 devices cited above in the devices database, the HIDS is able to establish an SSH connection to each device, one by one, and execute the verifications. In Table II we present the format expected of the outputs in the reports. A report is generated for each device and Table III summarizes the results.

The first interesting result regards the brute force technique that was used only against the Device 2, for which the Controller did not have the username and password. The proposed IDS discovered the username and password cisco and with this password the Controller continued to execute the verifications.

Then, as the common passwords of all users are checked, the IDS found in all 3 devices at least one user with a common password. The common passwords were also found to be leaked passwords, reinforcing the huge exposure that they represent.

The secret password was verified after that and again resulted in the default secret, which is cisco. Then the Controller looks for resources usage, analyzing the CPU and memory being used in our tested devices, concluding that there was no problem since none of them were using more than 80 percent of the total capacity.

A really important vulnerability that is surprisingly common in backbone components is the community string in the SNMP protocol being set as public for read and private for write operations in the MIB. An attacker with write permission access to a device can do massive damage, including turning off the device. In all 3 reports the warning to this vulnerability appeared, followed by the configuration of the SNMP protocol in each device.

Next, it was reported in the Device 1 that the IP CEF was not configured and the HTTP was enabled without HTTPS, representing a security vulnerability.

Finally, the open ports of the devices are presented in the reports. It is important to notice also that the HIDS found a CVE vulnerability for the device 1, as it is written in the description of the CVE that there is a possibility of DoS attack through Telnet. With that information and the information previously shown in the report that the Telnet is enabled and the port 23 is open, the administrator can take actions immediately.

In this validation process, the proposed HIDS set has proven to be capable of solving the problems presented in Section I related to common and default credentials. This distributed HIDS also provides key information relative to security breaches identified in generated reports for the administrator.

VI. Conclusion and Future Works

A distributed HIDS is proposed in this paper to execute a series of verifications and look for vulnerabilities in multiple devices in IoT backbones. A centralized station called Controller is used for storing and managing information regarding participant devices, being the main tool for the network administrator. The Controller interacts with a set of devices through SSH connections and for each device it runs verifications regarding vulnerabilities and security-related events.

Some basic vulnerabilities that are verified by the proposed system, like default or common usernames and passwords, are highly emphasized in this paper due to the large number of attacks that are taking advantage of them, especially by IoT botnets like Mirai.

By running tests in a local and controlled environment, it was possible to analyze three types of generated reports. The HIDS showed its ability to tackle the common username and password issue, either by executing tests directly in the devices for which the HIDS Controller had the necessary credentials, thus being able to find all weak passwords. Also, by using a brute force dictionary procedure on a device for which the Controller hadn't the credentials but managed to discover bad passwords, thus alerting the administrator of the possibility of brute force attacks to this device.

Based on other verifications regarding open ports, enabled protocols and information about the resources verified by the HIDS, the respective reports contained crucial information for the administrator of the network to take actions.

In future works, the distributed HIDS is going to be tested in a larger set of routers and switches from other different vendors. Also, considering the evolving relationship among vulnerabilities and attacks, a machine learning approach is under consideration for the analysis of the gathered information as a whole to allow the Controller to autonomously execute countermeasures against attacks.

Acknowledgment

The authors gratefully acknowledge the support from the IEEE VTS Centro-Norte Brasil Chapter, the Brazilian Research Councils CNPq (Grant 465741/2014-2 INCT on Cybersecurity), CAPES (Grant 23038.007604/2014-69 FORTE), and FAP-DF (Grants 0193.001366/2016 UIoT, 0193.001365/2016 SSDDC, and Call 01/2019), as well as the LATITUDE/UnB

TABLE III
RESULTS OF THE EXECUTED TESTS

Vulnerability	Device 1	Device 2	Device 3
Brute force	Not used	Used and successful. Username: cisco Password: cisco	Not used
Common password	1234, 1234	cisco	cisco, cisco
Leaked password	1234: 1256907	cisco: 5016	cisco: 5016
Common secret	Default secret	Default secret	Default secret
Memory utilization	0.13648%	0.09412%	0.28204%
CPU utilization over the last minute CPU utilization over the last 5 seconds CPU utilization over the last 5 minutes	21%, 54%, 9%	2%, 17%, 1%	2%, 2%, 1%
Community string in SNMP protocol	public RO, private RW	public RO, private RW	public RO
IP CEF	No	Yes	Yes
HTTP is enabled without HTTPS	Yes	No	No
Telnet	Open	Open	Closed
Port scan	22, 23, 80	22, 23	22
CVE Vulnerabilities	CVE-2000-0268. 2000-04-20T00:00:00. Cisco IOS 11.x and 12.x allows remote attackers to cause a denial of service by sending the ENVIRON option to the Telnet daemon before it is ready to accept it, which causes the system to reboot.	Not found	Not found

Laboratory (Grant 23106.099441/2016-43 SDN), the Ministry of the Economy (Grants 005/2016 DIPLA, 011/2016 SEST and 083/2016 ENAP), and the Institutional Security Office of the Presidency of the Republic of Brazil (Grant 002/2017).

REFERENCES

[1] Kaspersky. DDoS attacks in q1 2018. [Online]. Available: https://securelist.com/ddos-report-in-q1-2018/85373/

[2] M. Antonakakis, T. April, M. Bailey, M. Bernhard, E. Bursztein, J. Cochran, Z. Durumeric, J. A. Halderman, L. Invernizzi, M. Kallitsis, D. Kumar, C. Lever, Z. Ma, J. Mason, and D. Menscher, "Understanding the Mirai Botnet," in *Proceedings of the 26th USENIX Security Symposium*. Vancouver, BC, Canada: USENIX, 2017, pp. 1093–1110.

[3] S. Hilton. Dyn Analysis Summary Of Friday October 21 Attack. [Online]. Available: https://dyn.com/blog/dyn-analysis-summary-of-friday-october-21-attack/

[4] C. Kolias, G. Kambourakis, A. Stavrou, and J. Voas, "DDoS in the IoT: Mirai and other botnets," *Computer*, vol. 50, no. 7, pp. 80–84, 2017.

[5] "Annual CyberSecurity Report," Cisco 2018, Tech. Rep., 2018. [Online]. Available: https://www.cisco.com/c/dam/m/digital/elq-cmcglobal/witb/acr2018/acr2018final.pdf

[6] P. N. Raju, "State of the Art Intrusion Detection: Technologies, Challenges, and Evaluation," Master's Thesis, Linköping University, Linköping, Sweden, 2005.

[7] R. Gerhards, "The Syslog Protocol - RFC 5424," Internet Engineering Task Force (IETF), Tech. Rep., March 2009.

[8] H. G. C. Ferreira and R. T. de Sousa Junior, "Security analysis of a proposed internet of things middleware," *Cluster Computing*, vol. 20, no. 1, pp. 651–660, Mar 2017.

[9] A. Lazarevic, V. Kumar, and J. Srivastava, *Intrusion Detection: A Survey*. Boston, MA, USA: Springer US, 2005, pp. 19–78.

[10] A. Patel, Q. Qassim, and C. Wills, "A survey of intrusion detection and prevention systems," *Information Management & Computer Security*, vol. 18, no. 4, pp. 277–290, 2010.

[11] H.-J. Liao, C.-H. R. Lin, Y.-C. Lin, and K.-Y. Tung, "Intrusion detection system: A comprehensive review," *Journal of Network and Computer Applications*, vol. 36, no. 1, pp. 16 – 24, 2013.

[12] G. G. Helmer, J. S. K. Wong, V. Honavar, and L. Miller, "Intelligent Agents for Intrusion Detection," in *1998 IEEE Information Technology Conference, Information Environment for the Future (Cat. No.98EX228)*. Syracuse, NY, USA: IEEE, Sep. 1998, pp. 121–124.

[13] "Cisco Port Security," Cisco, Tech. Rep., 2018. [Online]. Available: https://www.cisco.com/c/en/us/td/docs/routers/7600/ios/15S/configuration/guide/7600_15_0s_book/port_sec.html

[14] "DHCP Snooping," Cisco, Tech. Rep., 2018. [Online]. Available: https://www.cisco.com/c/en/us/td/docs/switches/lan/catalyst6500/ios/12-2SX/configuration/guide/book/snoodhcp.html

[15] Y. F. Jou, F. Gong, C. Sargor, X. Wu, S. F. Wu, H. C. Chang, and F. Wang, "Design and implementation of scalable IDS for the protection of Network infraestructure," in *Proceedings DARPA Information Survivability Conference and Exposition. DISCEX'00*, vol. 2. Hilton Head, SC, USA: IEEE, Jan 2000, pp. 69–83 vol.2.

[16] G. Prashanth, V. Prashanth, P. Jayashree, and N. Srinivasan, "Using Random Forests for Network-based Anomaly detection at Active routers," in *2008 International Conference on Signal Processing, Communications and Networking*, Chennai, India, Jan 2008, pp. 93–96.

[17] S. R. Snapp, J. Brentano, G. V. Dias, T. L. Goan, L. T. Heberlein, C.-L. Ho, K. N. Levitt, B. Mukherjee, S. E. Smaha, T. Grance, D. M. Teal, and D. Mansur, "DIDS (Distributed Intrusion Detection System)-Motivation, Architecture, and An Early Prototype," in *Proceedings of the 14th National Computer Security Conference*. Washington, DC, USA: NIST, 1991, pp. 167–176.

[18] M. Silva, D. Lopes, and Z. Abdelouahab, "A Remote IDS Based on Multi-Agent Systems, Web Services and MDA," in *2006 International Conference on Software Engineering Advances (ICSEA'06)*, Tahiti, Tahiti, Oct 2006, pp. 64–64.

[19] B. V. Dutra, J. F. de Alencastro, F. L. de Caldas Filho, L. M. C. e Martins, R. T. de Sousa Júnior, and R. de O. Albuquerque, "HIDS by signature for embedded devices in IoT networks," in *Actas de las V Jornadas Nacionales de Investigación en Ciberseguridad (JNIC 2019)*. Cáceres, Spain: Universidad de Extremadura, Jun 2019, pp. 53–61.

Advances in Information Systems and Technology

IST 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—16th Conference on Advanced Information Technologies for Management
- DSH'19—1st Special Session on Data Science in Health
- InC2Eco'19—1st Workshop on Data Analysis and Computation for Digital Ecosystems
- ISM'19—14th Conference on Information Systems Management
- KAM'19—25th Conference on Knowledge Acquisition and Management

17^{th} Conference on Advanced Information Technologies for Management

WE are pleased to invite you to participate in the 17^{th} 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,** Wroclaw University of Science and Technology, Poland
- **Leyh, Christian,** University of Technology, Dresden, Germany
- **Ligęza, Antoni,** AGH University of Science and Technology, Poland
- **Ludwig, André,** Kühne Logistics University, Germany
- **Magoni, Damien,** University of Bordeaux – LaBRI, France
- **Michalak, Krzysztof,** Wroclaw University of Economics, Poland
- **Owoc, Mieczyslaw,** Wroclaw 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,** Wroclaw University of Economics, Poland
- **Stanek, Stanislaw,** General Tadeusz Kosciuszko Military Academy of Land Forces in Wroclaw, 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 fur Technik und Wirtschaft Berlin

- **Wolski, Waldemar,** University of Szczecin, Poland
- **Zanni-Merk, Cecilia,** INSA de Rouen, France
- **Ziemba, Ewa,** University of Economics in Katowice, Poland

Communication Papers of the Federated Conference on Computer Science and Information Systems pp. 93–100
DOI: 10.15439/2019F228
ISSN 2300-5963 ACSIS, Vol. 20

Efficient Production Monitoring on the Basis of Domain Ontologies by Utilizing IoT

Maha Riad
Faculty of Informatics and Computer Science
The British University in Egypt
Cairo, Egypt
riad_maha@hotmail.com

Amal Elgammal
Faculty of Computers and Information
Cairo University
Cairo, Egypt
a.elgammal@fci-cu-edu.eg

Doaa Elzanfaly
Faculty of Informatics and Computer Science
The British University in Egypt
Helwan University
Cairo, Egypt
doaa.elzanfaly@bue.edu.eg

***Abstract*—The Internet-of-Things (IoT) technologies and cyber-physical systems has facilitated production monitoring and control. However, researches and applications still lack a standardized framework and an integrated technological solution that can maximize the leverage of real-time monitoring. This can be achieved through enabling data transfer and exchange between all entities/organizations in supply chains and accordingly utilizing the monitored data. This paper introduces a framework for production monitoring that utilizes and integrates ontological model, which implements and integrates Semantic Sensor Network (SSN) ontology with production monitoring services. In addition, Complex Event Processing is integrated in the proposed model to enable event patterns identification and undertake the appropriate (pro-active) action accordingly. The framework is constructed based on ISA-95 and SCOR standards. The utility, applicability and efficacy of the proposed framework is validated by its application on a real-life large-scale case study in the domain of laser cutting machines.**

I. Introduction

IN manufacturing, it is crucial to ensure that production processes are ongoing without any problems. Accordingly, the manufacturing machines and used resources need regular monitoring to eliminate production losses in case of machines' unexpected breakdown or in case of producing final products with unacceptable quality. To avoid these problems traditional methods use human inspections [1] and planned regular machines maintenance [2]. However, Visual inspection is not sufficient; due to the limited detection range of observations [3]. In addition, relying on regular machine maintenance schedules might increase costs if machines are maintained prior to their need. On the contrary, it might not be detected that a machine needs earlier maintenance scheduling and would lead to machine break-down and significant losses.

The IoT technologies covers the previously mentioned gaps, through being able to monitor machines' conditions in a reliable and regular way [4]. As reported by Bera [5], one of the big industries that uses IoT devices is the 'manufacturing' industry; as it contributes by 40.2% from the different industries share in the IoT devices usage, in which the IoT devices are mainly used for robotic machinery control and fault diagnostics. In a survey carried out by Capgemini Digital Transformation Institute [6], it was found out that 57% out of 300+ companies had problems with not having IoT standards for data sharing and interoperability to follow. In order to overcome this challenge, semantic technologies must be used according to [7] ,[8] and [9].

Enabling real-time production monitoring and allowing direct communication with service providers or entities in the supply chain, would result in efficient production plans. Therefore, the aim of this paper is to introduce an integrated framework that is founded on formal ontological domain models and integrates IoT technologies (mainly sensors) with production monitoring services. Furthermore, the proposed model integrates complex event processing [10] to utilize data collected and identify abnormal event patterns arising during the production process. A real-life large-scale case study that is conducted in the context of the H2020 ICP4Life[1] project, is used to validate the usage, applicability and efficacy of the proposed framework and its implemented solutions.

The paper is divided as follows: In section (II), the background of IoT and ontologies is illustrated. In section (III), related work efforts are summarized. In section (IV), the used case study is demonstrated, while in section (V) the proposed framework is elaborated. Section (VI) includes implementation and evaluation details. Finally, section (VII) includes the conclusion and future work.

II. Background

In order to efficiently integrate the IoT technologies, it is crucial to take advantage of the semantic technologies capabilities, specifically ontologies. In this section, we will illustrate shortly the capabilities, components and structure of (A) IoT (B) Ontologies.

This research is partially funded by the EC Horizon 2020 project ICP4Life, under contract number FoF-05-2014 – 636862

† Corresponding author, amal@servtech.info

[1] ICP4Life: http://www.icp4life.eu/

©2019, PTI

A. Internet-of-Things

Internet-of-things (IoT) connects things/objects through the internet, through uniquely identifying each of the objects in the network and giving it the capability of communication with other objects/things. IoT architecture is composed of four main layers [4]: (i) sensor layer: which is concerned with all the physical devices involved to know the state of the object/thing, store its data or change its state. It includes sensors for collecting information about the object, as well as actuators for actuating on a specific object. It includes the RFID, that is responsible for storing data about the object. Additionally, this layer might include other devices used in data exchange and collection such as: smart phones, cameras, microphones, etc. (ii) network layer: which represents the means of data exchange throughout the network. (iii) service layer: This layer represents the means of services exchange (iv) interface layer: this layer represents the presentation/ displaying of the collected data to the end-users through different applications.

B. Ontologies and Semantic Sensor Network Ontology

As defined by Liu and Özsu [10], ontologies are used for the representation of knowledge in different domains by means of introducing the commonly used set of data or classes as well as their attributes. Moreover, they enable expressing the relationships between the different classes and their properties [10]. Ontologies are based on abstract representation of the data model, giving flexibility to separate the data models and the structures used or generally separating it from the implementation. Accordingly, the independence given to the data model in case of being used by an ontological model facilitates interoperability and standardization.

The Semantic Sensor Network (SSN) [11] is a standard ontology developed by W3C for representing sensors, their observations and their applied context. Moreover, it describes actuators, actuations and their applied context. It will play a crucial role in the proposed model as it represents a standard representation for the IoT used components in sensing and actuating and the main classes and properties needed.

III. Literature Review

A. Integration of Production Monitoring Systems and IoT

There are several models proposed by different researches and SW companies that utilize IoT in production monitoring. Each of the models has a specific perspective and facilitates certain system's functionalities as illustrated below and summarized in Table I:

Ding, Jiang and Su [12]proposed social manufacturing system that facilitates real-time production and transportation monitoring and analysis. The system was based on the use of RFID and was evaluated by a printing machinery company. The system framework is divided into three layers: (i) Physical layer: which mainly contains RFID tags, readers and antennas. (ii) Application layer: which connects the manufacturing company with the customers and suppliers by the aid of service-oriented architecture concepts. In the conceptual model they defined two types of databases that should be implemented: Private databases and Public database. Each of the supply chain entities should have a private database to store their related real-time information, while the public database is meant to store the common information between the different entities. (iii) Social layer: this layer integrates the system with different social media apps, such as Facebook and twitter, as the medium of communication.

Mi and Kara [13] introduced a methodology to formulate a real-time manufacturing monitor application based on two steps. First, they defined how to design the architecture of the system. Second, they defined how to choose the components. They tested their methodology on a small-scale case study, in which they monitored the environment temperature of an office. Their focus of assessment was mainly considered with the possibility and ability to use Wireless Sensor Networks (WSN) with IoT.

Ding and Jiang [14] started their paper with describing how an IoT-enabled job-shop is configured and how the production can be integrated with the IoT. Then they proposed a data model for event-driven data collection and analysis. Finally, they evaluated their work through a simulated case study.

Lee, Noh, Kim and Kang [15] proposed an architecture framework for Cyber-Physical Production Systems which aims for predicting quality and control of production processes. The framework integrates IoT, production processes and AI and they tested it on actual plants. The framework is mainly composed of three sub-systems: (i) Big Data Analytics system: which is responsible for storing the produced data from the IoT tools and the data manipulated by the manufacturing system. Also, it uses different machine learning tools to produce business insights (ii) Detection and Coordination System: this system is used for handling different events (iii) KPI Simulation system: Based on the collected real-time information and the production plan, a simulation model was carried out and produced reference KPIs for the expected quality and productivity issues.

Ahmed et al. [16] proposed a conceptual model that uses fact-driven statistical methodology to analyze machines' failure factors and causes and further facilitate critical failures prediction. Their conceptual model is divided to three layers: (i) Sensory Devices layer (ii) Data Storing and Sharing layer (iii) User Interface layer.

B. Ontological Models for Production/Manufacturing

From another perspective, there are several trials for proposing ontologies for production monitoring, which will be illustrated below.

Cao, Zanni-Merk and Reich [17] proposed an ontology for condition-monitoring based on ISO standards, claiming that it is more generalized than the current available condition-monitoring ontologies. They followed the iterative ontology development method, mentioned in [18], for developing the

TABLE I.
RELATED WORK MODELS AND TECHNOLOGIES USED

Paper	Area of Interest	IoT Technology	Means of Integration and Communication	Evaluation	Limitations
[12]	Real-time production & transportation monitoring	RFID	Social Media Service-Oriented Architecture	Printing Machinery Company	Each of the involved entities should have private database with the defined data interface of the model
[13]	methodology to formulate a real-time manufacturing monitor application	Sensors and RFID integrated with WSN	Cloud Platform	Office temperature environment monitoring	The case study focused on the limitations and abilities of usage of WSN and RFID in the manufacturing company and did not proposed a standardized interoperable model that would allow data exchange with other entities/organizations such a suppliers and customers
[14]	IoT-enabled smart job-shop production	RFID antennas readers, and active tags	None	Job-shop production lab simulation	The gathered data is utilized inside the production company only
[15]	Cyber-Physical production systems monitoring and control	General tools used not mentioned specifically	None	Piston engine factory	The gathered data is utilized inside the production company only
[16]	Production critical failure analysis	Sensory Devices	Cloud Platform	Beverage production company	Lack of standardized approach for the exchanged data manipulation by the different organizations.
[17]	Condition monitoring ontology	Sensors	Ontology	ISO standards	The classes used in the ontology has limited scope and do not fully integrated the IoT, Production processes and collected data utilization
[19]	Smart product service ontology	Sensors	Cloud infrastructure & Ontology	Laser cutting machine manufacturer	The classes used in the ontology has limited scope regarding the data collected utilization and the IoT classes
[20]	Intelligent conditional monitoring ontology	-	Ontology	Rotating machinery	The classes used in the ontology has limited scope and do not fully integrated collected data utilization and IoT

ontology. Its core classes include: System, State, Process, Parameter, Sensor, Fault, Failure and Behavior.

In an earlier work related to ICP4Life project, Maleki *et al.* [19] presented a framework that enables products' customizing services by the aid of sensor ontology and this was validated by using an industrial use case.

Cao *et al.* [20] proposed an intelligent conditional monitoring ontology. It includes the main classes of production processes and faults handling, but it ignores classes related to the IoT technology.

As a consequence of checking the available related work and the proposed models, it was found that none of them provide an interoperable standardized solution that can allow knowledge sharing through different organizations. Moreover, those models do not allow subsequent utilization of the gathered and exchanged data through using analytics or event patterns identification.

IV. CASE STUDY

The case study is carried out for turbine engine manufacturer, which is in our context the entity that needs production monitoring for laser cutting machines used in production. In related work [21] of the H2020 ICP4Life project, the objectives were concerned with co-designing and specifying the machine components and design according to the user's (turbine engine manufacturer) requirements. However, in the scope of the paper context, the objective is to utilize sensors added to the machine and to ensure data unification, to facilitate and enable data exchange with the service provider.

The user uses a web application to customize the product's sensors related to the required after-sales services. Furthermore, the user monitors the information collected by the sensor using dynamic charts and is altered through the system in case of exceeding a primary determined threshold for the sensor readings, as shown in figure 2. For instance, the user can choose the temperature sensor for the monitoring and control service by dragging a 2-D image that represents the sensor and service in addition to the main components used in the laser cutting machine composition as shown in figure1. When the machine is working, the temperature changes recorded by the sensor will be reflected on the screen as shown in figure2.

The manufacturing blueprint concept introduced by [22] is used to ensure having a smart manufacturing network through cross-functional integration and communication.

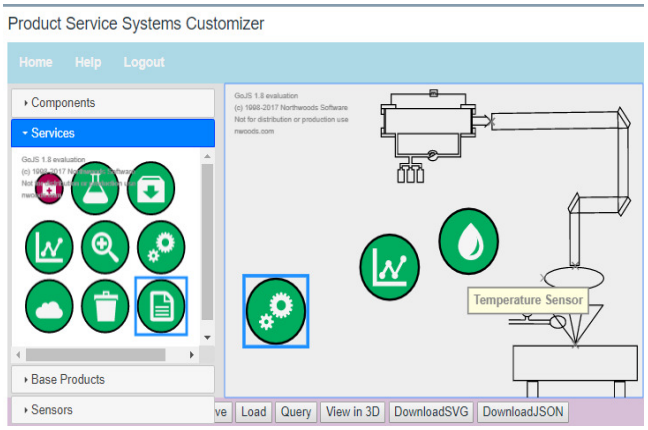


Fig. 1 Product Service Systems Customized after Sensors customization addition to the solution proposed in [23]

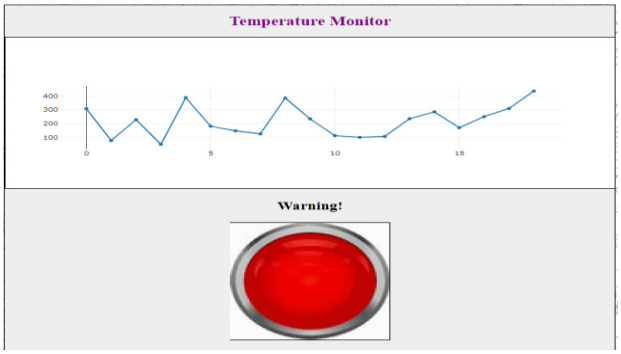


Fig. 2 Temperature Monitoring and Notification

V. Proposed Framework

Reviewing the main modules involved in product-service systems (PSS) stated by [23], which are: (i) PSS requirements elicitation. (ii) PSS configuration and customization (ii) Production planning (iii) Production execution. (iv) Production Monitoring, we will capitalize our proposed work in this paper on the production monitoring requirements.

The main objective is to propose a reliable framework that enables real-time production monitoring and allows direct communication with the supply chain entities. Accordingly, the proposed framework is composed of four layers to meet the required objective as shown in figure 3.

A. Data Collection Layer

The first layer in the framework, is the 'Data Collection Layer' and it includes all the tools used to collect the data related to the machines' conditions and to the attributes that affect the product quality. It mainly contains all the IoT hardware such as sensors, readers, cameras, microphones. The needed hardware is determined according to the monitoring criteria. They are mainly installed in the machine context and their data is shared between the manufacturer and service provider. For example, in the case study temperature sensors where added to the cutting laser machine to measure the environmental conditions.

B. Data Storage & Management Layer

The data storage layer ensures two aspects: The interoperability and the data sharing. Storing the data in a data center that is accessible by all interested parties would facilitate communication. However, exchanging data that is meaningful and understandable needs a standardized approach. Accordingly, to unify data, the data storage & management layer are based on ontology or the blueprints concept [22].

There are three main levels of ontology that form a hierarchy: (i) The Top-level ontology: which It defines the general terms, entities and their relationships. The main feature of top-ontology is that it can be implemented in several domains e.g. Semantic Sensor Network (SSN) ontology, as it includes the main classes related to the sensor network without defining the context or domain that they will be integrated in. (ii) The domain-level ontology: It defines entities, attributes and relationships related to a specific domain, which in our case will be the cutting-laser machines domain. The most common sensors involved in the observation of cutting laser machines will be added as sub-classes of the sensor class. (iii) The application-level ontology: It contains the actual instances instantiated from the domain ontology in the running application. In the context of the case study, the actual sensor used for monitoring temperature is added as instance from the temperature sensor class.

Top-Ontology/Blueprints:

[22] introduced five main blueprints related to the manufacturing scope:

- Partner/Stakeholder Blueprint: It is mainly concerned with the knowledge related to the partner's company, its details and capabilities.
- Product Blueprint: It contains knowledge related to the product, its materials, needed resources, components and properties.
- Product-Service: It focuses on the after-sale services needed for the product such as machine maintenance and upgrades.
- Production-Plan Blueprint: It describes the process flow, involved activities and resources for actual production.
- Quality Assurance Blueprint: KPIs metrics and the knowledge needed to ensure that the product's and machine's quality are reflected in this blueprint.

In the context of our framework objectives, four main ontologies are integrated to the SSN ontology to reach the objective of the proposed framework.

- Service Blueprint: The product service blueprint proposed by [22] is extended to integrate the appropriate classes from the SSN, which facilitates providing related services such as production monitoring and control. Primarily, the customer determines the needed services for their resources

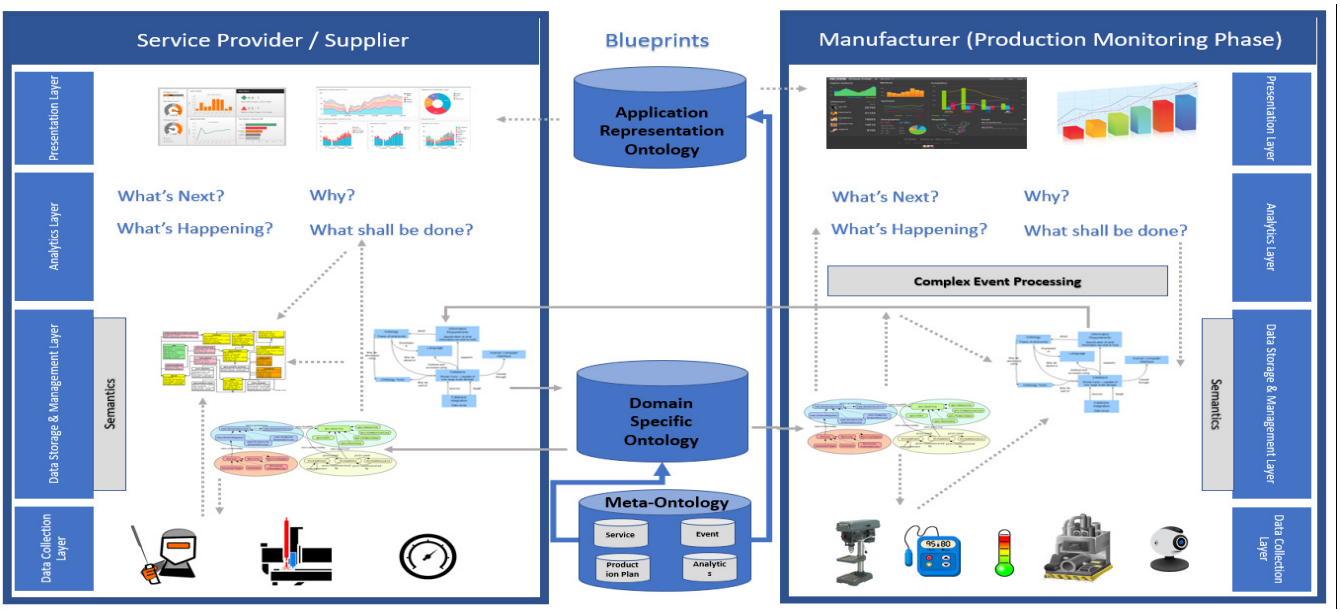


Fig. 3 Proposed Framework

(products) and then accordingly, the sensors are determined. Finally, the observation process takes place. Some examples of the classes included in the blueprint are: (i) Service: It describes the service. (ii) Service provider: It represents the entity that provide the service. (iii) Customer: It represents the customer who requests the customized service. (iv) Sensor: It is one of the main classes integrated from the SSN as the sensor determination will be determined based on the service.

- Event Blueprint: It represents the entities used in detecting and monitoring events. Event tracking is facilitated after the integration of sensors, as they enable collecting real-time data and identifying abnormal patterns. An example of the classes included in the blueprint: (i) Event class: It represents the event. (ii) Listener class: It represents the class that listens to the result and act accordingly by comparing the readings to the thresholds. (iii) Event handler class: It acts upon managing how the event and its readings are managed with subscribers or listeners
- Production Blueprint: It describes the production processes and their segments. It also describes the needed resources and production processes monitoring rules. Examples of included classes: (i) Process Segment: Which includes the main steps carried out to perform specific process. (ii) Production rules: Which states the constrains and rules related to the production process. (iii) Product segment: It is the overlap definition between the production rules and the bill of resources (required resources) which are defined in the segment requirements. (iv) Resources: Which are the resources involved in the production process.
- Analytics Blueprint: It represents the classes needed for analyzing the collected data and results. In the current scope, it only includes the possible actions that could be taken after a specific event pattern detection. There are three main types of action: (i) Descriptive action: Which is just descriptive for the situation in the form of alert or message notification with the result. (ii) Predictive action: Which applies predictive model on the collected data to reach specific insight. (iii) Actuator action: Which represents the action that should be carried out by an actuator.
- Production Monitoring Blueprint: It represents the means of integration and relationships representation between the main classes used from each of the previous blueprints. The blueprint is shown in figure 4 and is demonstrated shortly in the upcoming paragraph.

Using the product production rules, the monitoring rules are determined, which describes the needed observable property and defines the needed listeners based on the threshold. The listeners listen to the event results of the observations made by the sensors. If the listener detects a specified abnormal pattern of events, an appropriate action is triggered. The action can either be informative, in which it describes what has happened in terms of sending warning or notification, or the action can be actuatable action, which is an action that triggers a specific actuator to act on specific actuatable property, or the action can be predictive action.

An example of the instantiated individuals from the blueprint in the case study is as follows:

The monitoring rule specify that the temperature cannot exceed 40 degree Celsius and there should

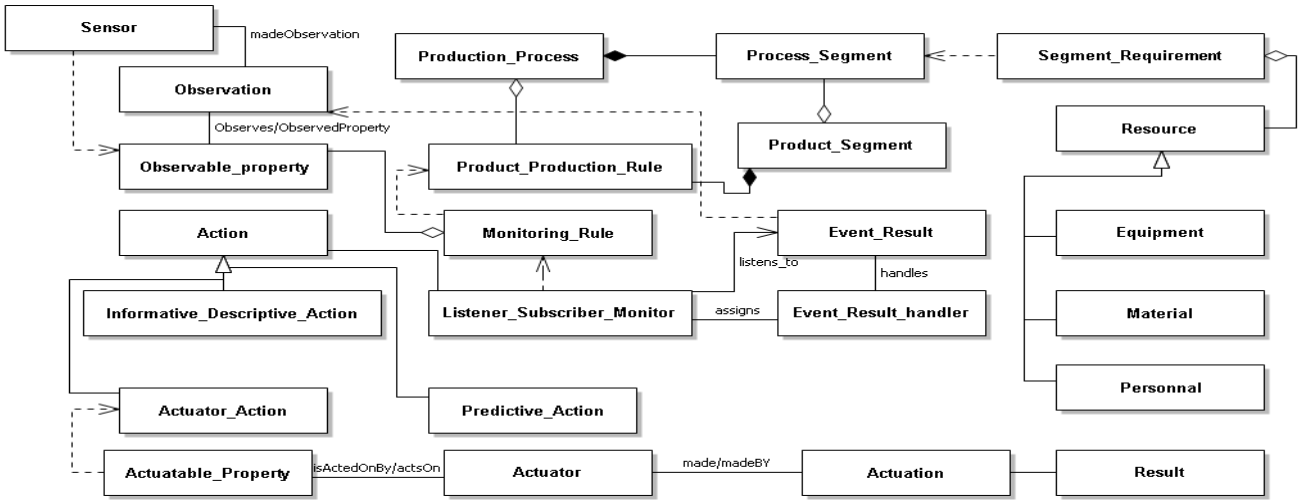


Fig. 4 Production Monitoring Blueprint

not be increase in temperature in 3 consequent readings by more than 0.1%. Accordingly, it is determined that machine temperature is set as observable property and temperature sensor is needed. The observations' results will be listened to by the listener to detect any matching pattern with the threshold and restrictions. In case of pattern detection, an informative action is taken through sending warning on the application screen.

Domain Specific Ontology:

The classes are customized depending on the domain context. For example: In case of the context of cutting laser machine, the cutting laser machine class will extend class equipment.

Application Representation Ontology:

Introducing a domain-ontology for a specific context like the laser cutting machines allows interoperability and knowledge sharing between different entities. However, if we need to map the process that sends data to the front-end dynamically, each organization should develop its own customization terms, to link the back-end data to the front-end display. To facilitate this process, we propose the 'Application Representation Ontology'. The Application Representation Ontology is responsible for defining how the instances created shall be displayed. For example, to display the temperature sensor, it can include the displaying picture, location and dimensions of picture on the screen, the color of the surrounding frame and so on. This would allow the instances and their display properties to be read directly from the blueprints and displayed to the end-user.

Data Exchange between different entities using blueprints:

After the data is collected (at the manufacturer side) with the IoT tools the data is mapped to the domain ontology and the instances are created based on the data, the relative instances are sent to (the Service Provider). At the Service Provider the instances are compared against the domain Specific ontology and save its instances accordingly.

A. Complex Event Processing

As illustrated by [10], complex event processing is about detecting event patterns in real-time through monitoring the input data and listening to it. Then, in case of matching a pattern sequence an appropriate action is taken. It is used in the model as a main component for detecting event patters and acting correspondingly.

B. Analytical Layer

The analytical layer utilizes the collected data through demonstrating and justifying what has happenedpredicting what will happen and can also recommend the possible proactive action. This is reached by implementing descriptive, diagnostic, predictive and prescriptive analytics.

C. Presentation Layer

The presentation layer is the gateway for presenting the collected data and gained knowledge to the end users (stakeholders). It may include dashboards, managerial reports and user-friendly GUIs for data representation and production system control.

VI. IMPLEMENTATION AND EVALUATION

To ensure the reliability of our model: (i) Related standards were considered, which are ISA-95 and SCOR (ii) A standard ontology to represent the IOT classes integration was used, which is the SSN proposed by W3C. (iii) Limitations detected in other ontologies proposed by researches were overcame. (iv) The guidelines of how to create ontologies and the carried-out steps stated by [18] were followed. (v) The case study illustrated in section IV was carried in manufacturing cutting laser machines domain.

In the case study, the following procedures were carried out during implementation:

A. Creating Instances

Instances of the ontology/blueprint were created and stored using 'Protégé' ontology editor [24]. Example of created instances are shown in Table II and Figure 5:

TABLE II.
CREATED INSTANCES EXAMPLE

Meta-Ontology Level Class	Domain Level Class	Instantiated Individual
Product	Cutting Laser Machine	Cutting Laser Machine Type A
Product Component	Lens	Lens X
Event	Temperature Event	Temperature Event A
Product Segment	Cutting Laser Machine Product Segment	Co2 Cutting Laser Machine Product Segment
Sensor	Temperature Sensor	Temperature Sensor A

B. Ability of Reading and Writing Instances (Manipulating Blueprints)

As the target of the blueprints is to have a common knowledge model that enables data unification and exchange, creating instances from blueprints as well as manipulating them is essential. Reading and writing from the blueprints is carried out by the aid of Jena, a java library, which enables reading/writing from/in ontology files.

However, in order to make the process self-dependent of accessing blueprints until displaying content of the system. We recommend having a fourth ontological layer, above the third layer of instances to contain the main ontological classes needed in the display of the instances. For example, the GUI data properties that are specified for the temperature sensor includes the filling shape, the location at the screen, the color of the displayed picture, etc.

C. Used Scenario and Instances for Simulation

Using the Monitoring rules, it is determined that the temperature event of the surrounding environment of the machine needs to be monitored. Accordingly, the observable property is determined to be the temperature. Moreover, the temperature sensor is used for the observation. The listener listens to the results and according to the production rules, it compares the results pattern to the identified threshold. When the listener identifies abnormal patterns, an appropriate action is taken. In our system we have simulated the 'informative_descriptive_action' and a warning notification is sent. The prototype was implemented using a web app developed on Eclipse Jee Mars, version (4.5.2). Tomcat Server V8.0 was used. To handle back-end and front-end synchronization, AJAX was used aside from Jsp, java, HTML and JavaScript.

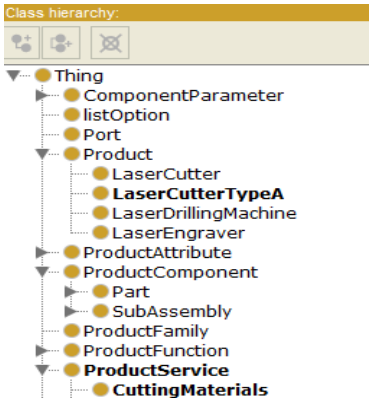


Fig. 5 classes in Protégé

D. Discussion

The proposed model and system were able to provide an ontological model for production monitoring and mainly cutting laser machines domain. The data was saved and retrieved with reference to the model. The system enabled utilizing the data collected through complex event processing, which identifies abnormal patterns. Accordingly, the system was able to cover the gaps found in the other related works as seen in Table III.

TABLE III.
COMPARISON BETWEEN THE PREVIOUS MODELS AND THE PROPOSED MODEL

Paper	Monitoring using IoT technologies	Means of Communi-cation & Data Exchange	Interop-erability	Utilizing collected data / Analysis
[12]	Yes	Yes	No	Yes
[13]	Yes	Yes	No	No
[14]	Yes	No	No	Yes
[15]	Yes	No	No	Yes
[16]	Yes	Yes	No	Yes
[17]	Yes	Yes	Yes	No
[19]	Yes	Yes	Yes	No
[10]	No	Yes	Yes	Yes
Proposed Model	Yes	Yes	Yes	Yes

VII. CONCLUSION AND FUTURE WORK

This paper aims to provide efficient framework for production monitoring based on domain ontologies and IoT utilization. Accordingly, the proposed contribution of the paper is threefold: (i) It defines the main layers needed for an integrated framework that enables real-time monitoring through using IoT technologies. In addition, it ensures data unification through an ontological model for data exchange. Moreover, it utilizes the collected data through integrating

complex event-processing for abnormal events detection and supports the monitoring process via a presentable GUI. (ii) It proposes an ontological model for production monitoring based on the blueprints' concepts introduced in [22] and the ISA-95 standards [25]. (iii) It introduces 'Application Representation Ontology' as a fourth level of ontologies mapping to define the means of displaying ontology's individuals in applications and systems. A case study was applied in the cutting laser machines domain for framework applicability evaluation.

Current and future work efforts are in diverse directions related to Manufacturing and Blueprints Concepts. However, a main focus for the consequent work related to the proposed work in this paper, is the utilization of real-time data using analytics by the aid of ontology reasoning. Business Analytics will not only facilitate predicting crucial production concepts, but also will enable reaching a proactive approach that will make the production leaner. In another direction, the blueprints are to-be enhanced and extended in an iterative manner, to make use of industry 4.0 capabilities and technologies.

VIII. ACKNOWLEDGEMENT

We wish to thank PRIMA INDUSTRIE S.p.A. (https://www.primaindustrie.com/) for providing the essentials of the pilot study in this paper. Special thanks go to Dr. Indika Priyantha, Jheronimus Academy of Data Science, the NL, for contributing to the development of the tool-suite. The authors are grateful to engineers José Luis Ajuria and Ion Ruiz at SEGULA Tecnologias España for their contributions to the 3D CAD/CAM capabilities of the tool-suite.

REFERENCES

[1] P. Foster, "Manufacturing Quality Control: The Difference Between Product and Process Audits," *Beacon Quality*, 2018. [Online]. Available: https://www.beaconquality.com/blog/manufacturing-quality-control-the-difference-between-product-and-process-audits. [Accessed: 23-Apr-2019].

[2] M. Ardolino *et al.*, "The role of digital technologies for the service transformation of industrial companies," *Int. J. Prod. Res.*, vol. 7543, no. May, 2017. DOI: https://doi.org/10.1080/00207543.2017.1324224

[3] V. Chaudhary, I. R. Dave, and K. P. Upla, "Automatic visual inspection of printed circuit board for defect detection and classification Automatic Visual Inspection of Printed Circuit Board for Defect Detection and Classification," in *2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, 2017. DOI: https://doi.org/10.1109/WiSPNET.2017.8299858

[4] M. Riad, A. Elgammal, and D. ElZanfaly, "Efficient Management of Perishable Inventory by Utilizing IoT," in *2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, 2018. DOI: https://doi.org/10.1109/ICE.2018.8436267

[5] A. Bera, "80 Mind-Blowing IoT Statistics (Infographic)," 2019. [Online]. Available: https://safeatlast.co/blog/iot-statistics/. [Accessed: 25-May-2019].

[6] Capgemini D. T. Institute, "Unlocking the business value of IoT in operations," 2018.

[7] S. Dupont, A. Achour, L. Deru, and N. Matskanis, "Bringing Dynamics to IoT Services with Cloud and Semantic Technologies An Innovative Approach for Enhancing IoT based Services," in *IoTBD 2016 - International Conference on Internet of Things and Big Data*, 2016, pp. 185–190. DOI: https://doi.org/10.5220/0005933001850190

[8] I. Szilagyi and P. Wira, "Ontologies and Semantic Web for the Internet of Things – A Survey," in *42nd IEEE Industrial Electronics Conference*, 2016. DOI: https://doi.org/10.1109/IECON.2016.7793744

[9] M. Ganzha, M. Paprzycki, W. Pawłowski, P. Szmeja, and K. Wasielewska, "Semantic technologies for the IoT – an Inter-IoT perspective," in *2016 IEEE First International Conference on Internet-of-Things Design and Implementation (IoTDI)*, 2016. DOI: https://doi.org/10.1109/IoTDI.2015.22

[10] T. Rohrmann, "Introducing Complex Event Processing (CEP) with Apache Flink." 2016.

[11] W3C, "Semantic Sensor Network Ontology." Spatial Data on the Web Working Group, 2017.

[12] K. Ding, P. Jiang, and S. Su, "RFID-enabled social manufacturing system for inter-enterprise monitoring and dispatching of integrated production and transportation tasks," *Robot. Comput. Manuf.*, vol. 49, no. 2018, pp. 120–133, 2017. DOI: https://doi.org/10.1016/j.rcim.2017.06.009

[13] W. Li and S. Kara, "Methodology for Monitoring Manufacturing Environment by Using Wireless Sensor Networks (WSN) and the Internet of Things (IoT)," in *The 24th CIRP Conference on Life Cycle Engineering Methodology*, 2017, vol. 61, pp. 323–328. DOI: http://dx.doi.org/10.1016/j.procir.2016.11.182

[14] Kai Ding and P. Jiang, "RFID-based Production Data Analysis in an IoT-enabled Smart Job-shop," *IEEE/CAA J. Autom. Sin.*, vol. 5, no. 1, pp. 1–11, 2018. DOI: https://doi.org/10.1109/JAS.2017.7510418

[15] J. Lee, S. Do Noh, H. Kim, and Y. Kang, "Implementation of Cyber-Physical Production Systems for Quality Prediction and Operation Control," *Sensors*, vol. 18, no. 1428, 2018. DOI: 10.3390/s18051428

[16] S. Ahmad, A. Badwelan, A. M. Ghaleb, A. Qamhan, and M. Sharaf, "Analyzing Critical Failures in a Production Process : Is Industrial IoT the Solution ?," *Wirel. Commun. Mob. Comput.*, vol. 2018, 2018. DOI: https://doi.org/10.1155/2018/6951318

[17] Q. Cao, C. Reich, and P. Afonso, "Towards a Core Ontology for Condition Monitoring," in *Procedia Manufacturing*, 2019, vol. 28, pp. 177–182. DOI: https://doi.org/10.1016/j.promfg.2018.12.029

[18] N. F. Noy and Deborah L. McGuinness, "Ontology Development 101: A Guide to Creating Your First Ontology." [Online]. Available: https://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html. [Accessed: 25-Apr-2019].

[19] E. Maleki *et al.*, "Ontology-based framework enabling smart Product-Service Systems : Application of sensing systems for machine health monitoring," *IEEE Internet Things J.*, vol. 2327, no. 4662, 2018. DOI: https://doi.org/10.1109/JIOT.2018.2831279

[20] Q. Cao, F. Giustozzi, C. Zanni-merk, F. D. B. De, and C. Reich, "Smart Condition Monitoring for Industry 4 . 0 Manufacturing Processes : An Ontology-Based Approach," *Cybern. Syst. an Int. J.*, 2019. DOI: https://doi.org/10.1080/01969722.2019.1565118

[21] D. Bogataj, M. Bogataj, and D. Hudoklin, "Mitigating risks of perishable products in the cyber-physical systems based on the extended MRP model," *Int. J. Prod. Econ.*, vol. 193, no. April 2016, pp. 51–62, 2017. DOI: https://doi.org/10.1016/j.ijpe.2017.06.028

[22] M. P. Papazoglou and A. Elgammal, "The Manufacturing Blueprint Environment : Bringing Intelligence into Manufacturing," in *23rd ICE/IEEE International Technology Management Conference*, 2017, pp. 772–781. DOI: 10.1109/ICE.2017.8279960

[23] M. P. Papazoglou, A. Elgammal, and B. J. Krämer, "Collaborative on-demand Product-Service Systems customization," *CIRP J. Manuf. Sci. Technol.*, 2018. DOI: https://doi.org/10.1016/j.cirpj.2018.08.003

[24] BMIR, "Protégé." [Online]. Available: https://protege.stanford.edu/.

[25] ISA, "ISA–95.00.01–CDV3 Enterprise-Control," pp. 1–165, 2008.

Communication Papers of the Federated Conference on Computer Science and Information Systems pp. 101–106
DOI: 10.15439/2019F299
ISSN 2300-5963 ACSIS, Vol. 20

A mixed-methods measurement and evaluation methodology for mobile application usability studies

Paweł Weichbroth
WSB University in Gdansk
Grunwaldzka 238A,82-266 Gdansk, Poland
Email: pweichbroth@wsb.gda.pl

***Abstract*—Low usability of mobile application is thought to diminish the perceived level of the quality by a user whose experiences substantially determines its market success or failure. However, while a single method of measurement employed to study usability may produce an unreliable or incomplete evaluation outcome, in this paper, contrarily we propose to take advantage of both qualitative and quantitative methods adequate to collect data, that would describe all usability attributes. In particular, in the scope of mobile application usability studies, this paper (*i*) depicts the main assumptions of elaborated M4MAUME methodology, (*ii*) describes the self-developed software tool (RVDA) for retrospective video data analysis, (*iii*) specifies the experimental setup, and (*iv*) discusses the preliminary results obtained from 20 experiments performed on four different groups of mobile applications. Eventually our findings lead us to the conclusion that mixing different methods have produced reliable and valuable outcomes which may be used to improve and manage usability in current and future projects, as well as to enhance existing software quality assurance (SQA) programs.**

I. Introduction

USABILITY of desktop interfaces has been widely studied. During the past 30 years, the human-computer interaction (HCI) community has developed a plethora of concepts [1,2,3], inspection techniques [4,5,6], frameworks [7,8,9], and heuristics [10,11,12] in order to better understand, measure and evaluate usability, which simply aims at eventually delivering better software products [13].

The long lived marriage of hardware and communication technologies has brought about the inevitable shift from desktop to mobile computing [14,15], leading to new user requirements regarding mobile applications [16,17,18]. As a consequence, usability engineers face the issue of usability measurement and evaluation in the instance of new software settings, including user interface (UI) design [19], connectivity [20], context-awareness [21], as well as hardware capabilities, concerning screen size [22], storage space [23] and overall performance [24].

In light of systematic literature review and analysis, covering the volume of 791 documents, indexed by Scopus database and published between 2001 and 2018, we determined the research methods applied to measure and evaluate usability of mobile applications [25]. Our findings also show that in only a few studies, collecting data concerned more than one technique, since very few utilized the retrospective video data analysis as well. This research gap inspired us to establish a new laboratory, equipped with both hardware apparatus and software tools, as well as elaborate congruent methodology with all the necessary data collection techniques.

The remainder of the paper is organized as follows. In Section II, a brief research background is described, followed by the M4MAUME methodology (Section III). In Section IV the RVDA tool is specified. In Section V we detail the general settings of the experimental setup, followed by a discussion of the preliminary results obtained from the undertaken experiments. Finally, Section VII concludes the paper.

II. Definitions, Measures, and Methods

The generally accepted definition of usability is the one, given by ISO in 92411-11, which states that usability is 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" [26]. This definition has been adopted in the majority of studies in the subject of mobile application usability.

In our latest study [25] we pointed out the ten most frequent usability attributes, namely: efficiency (70%), satisfaction (66%) and effectiveness (58%), learnability (45%), memorability (23%), cognitive load (19%), errors (17%), simplicity (13%), ease of use (9%) and navigation (8%), where the percentage in brackets show the relative frequency of occurrence of the attribute in reference to all extracted documents.

Efficiency is the ability of a user to complete a task with speed and accuracy [27]. Efficiency is measured in a number of ways, such as the duration to complete a given task, or the duration to finish a set of tasks [27,28]. In general, two methods are put into use: controlled observation, and survey.

Satisfaction is a user's perceived level of comfort and pleasure, or a user's perceived level of fulfillment of their expectations and needs [29]. Satisfaction is measured by using questionnaires and other qualitative techniques, typically used to capture a user's intangible attitude towards an application [29,30].

Effectiveness is the ability of a user to complete a task in a given context [31]. It is measured by the number of successfully completed tasks, the number of steps required to complete a task, the number of double taps unrelated to the

©2019, PTI

operation of an application, and the number of times that a back button is used on the mobile device [31,32]. To collect all necessary data to estimate measures, two methods are in common use: controlled observation, and survey.

Learnability is defined twofold: first-time and over-time. The former refers to the degree of ease with which a user can interact with a newly-encountered application without seeking guidance or referring to documentation. It is measured by the number of attempts to solve a task, the number of assists when performing a task, and the number of errors performed by a user [33]. Contrarily, the latter is the capacity of a user to achieve proficiency with an application. Typically, a user's performance during a series of tasks is observed to measure how long it takes these participants to reach a pre-specified level of proficiency [33,34].

Memorability is the degree of ease with which a user can remember how to use an application effectively. It is measured by asking users to perform a series of tasks after having become proficient with the use of the application, and afterwards asking them to perform similar tasks after a period of inactivity. To determine how memorable the application was, a comparison is made between the two sets of results [35,36]. In a few cases, eye tracking techniques were utilized to a greater extent in the usability studies.

Cognitive load refers to the amount of mental activity imposed on a user's working memory during application usage [37]. Cognitive load theory differentiates cognitive load into three types: extraneous, intrinsic and germane [38]. Extraneous refers to instructional and presentation schemas, caused by the mental activities and elements that do not directly support application usage; intrinsic refers to the task complexity, caused by the number of elements in a task and the degree to which these elements are related to each other; germane regards to the amount of mental effort used to form schemas and actively integrate new information with prior knowledge during application usage. In the practice of cognitive load measurement, instruments such as a subjective rating scale, a thinking aloud dual task protocol or eye tracking have been used [32,39].

Errors relates to the amount and type of errors which occur during task performance by a user [40]. On the other hand, it is the ability of an application to recover from occurred errors. Both these definitions also respectively reflect a way the attribute is measured [40,41]. Controlled observation and survey are the only methods used to observe both application and users' performance or to collect a users' perceived level of application usage without error.

Simplicity is the degree of being easy to understand or being uncomplicated in form or design, described by such characteristics as the number of menu levels, the number of performed gestures to reach a destination object, and the duration of searching a button to perform a specific function. On the other hand, simplicity is the level of comfort with which a user is able to complete a task, measured by predefined statements with the Likert-scale rating.

Ease of use is the perceived level of the user's effort related to usage of the application. The survey instrument is used to collect data from users on perceptions concerning their experienced interaction with the application [40,42].

Navigation is the perceived level of the user's effort to access relevant information. Similarly, the survey instrument is applied to collect data from users on perceptions concerning their understanding of the information architecture [43,44,45].

III. THE 3M4MAUME METHODOLOGY

Theoretical background

By definition, methodology is a body of methods, rules and postulates employed by a discipline [46]. In brief, a method is a process of doing something [47], a rule is a prescribed guide for conduct [48] and postulate is a hypothesis advanced as an essential premise of a train of reasoning [49].

Theoretical triangulation is the use of multiple theories or hypotheses when examining a phenomenon [50], while methodical triangulation has been defined as multimethod, methods triangulation, or simply mixed-method [51]. The application of triangulation, in the scope of multiple sources of data, enhances the reliability of results [52]. There is a direct link between data triangulation and data saturation, where the former is a method to establish the latter.

The quantitative paradigm is based on positivism [53], where evidence is characterized by empirical research. This group of methods emphasize objective measurements and the numerical analysis of data collected through polls, surveys and questionnaires. In contrast, the qualitative paradigm is based on interpretivism [54] and constructivism [55]. The three most common methods are: focus groups (group discussions), individual in-depth interviews, and participant observation [56].

Verbal protocol analysis (VPA) is a method for collecting and analyzing verbal data regarding cognitive processing [57]. In other words, VPA is the record of spoken thoughts, provided by subject when thinking aloud during, or immediately after completing a task [58]. It is based on the premise that in order to capture a participants real and authentic experiences, we must allow them to express themselves freely [59]. Alternatively, in a narrow sense, this method (protocol) is also defined as think aloud or thinking aloud, and is relatively free-form [60] and open-ended.

Main assumptions

The Mixed-Methods Methodology for Mobile Application Usability Measurement and Evaluation (3M4MAUME) is a body of three integrated methods, namely: (1) survey, (2) participant observation and (3) verbal protocol analysis. In this regard, this multi-faceted approach is designed to take advantage of both quantitative (1) and qualitative (2 & 3) approaches, which in a specified combination are able to

provide a full panorama on an application's quality-in-use characteristics, as well as user's attitudes and perceptions.

In particular, each survey is implemented according to a series of steps, each of which includes an application-specific functionality and defined set of formats and procedures.

Participant observation simultaneously takes place in conjunction with thinking aloud protocol. To preserve privacy, and encourage subject confidence, only an observer is present during the testing session. Moreover, each participant can ask questions to clarify aspects of the task, but all answers given are as brief as possible to minimize the time burden during the application testing session.

The observer has access to an external monitor which displays the participant's interaction with the application in real-time mode, while a third-party application records the session, including both video and voice data.

After the test, the participant is asked to run through the ideas generated and explain the thinking behind them, where not already mentioned.

Task no 1. Data collection

The usability testing procedure aims to collect data for analysis, and is the process described by a fixed sequence of three steps:

1) Pre-testing questionnaire collects the demographic data and ten statements regarding a mobile applications usage and observed usability issues, and following the self-assessment of the skills used to perform five tasks which correspond to the main application functionality, as well as to the tasks assumed to be performed afterwards by the participant.
2) Application testing session is described by the protocol which assumes participant recording during application usage, coded by the audio/video hardware apparatus in order to collect both voice and video data, without assistance or guidance. Moreover, participants are always encouraged to think aloud about the application properties and behavior, as well as to speak frankly of any other important issues.
3) Post-testing questionnaire which aims to reproduce the perceived quality in use, specified by ten usability attributes where each is typified by at least five statements.

The pre and post questionnaires are individually administered to the participants, before and after the test respectively. Therefore, the output of this task concerns two questionnaires, and an audio-video binary file, which gathered from the whole group, are afterwards verified, and eventually serve as the input for the data analysis task.

Task no 2. Data analysis

The data analysis task is the process described by a fixed sequence of the following three steps:

1) Inspecting video content that comprises annotation procedures in which the user's actions and application responses are identified, separated and marked on the timeline.
2) Documenting all identified application bugs, defects, errors, and any reported usability issue.
3) Extracting numerical values required to calculate usability attribute measures.

It is worth noting that to perform the above task, specific software tools are required to obtain reliable outcomes with acceptable accuracy. Moreover, all necessary calculations are undertaken, regarding quantifying attributes measures and estimating structural model parameters.

Task no. 3. Information visualization

The third stage involves information visualization on the dashboard to empower cognition of the extracted and analyzed data which precedes and facilitates usability evaluation. For that purpose we specified a weighted and labeled graph. Vertices show the sequence of user's actions, and weights are used to represent the duration of application responses. In other words, the graph is a reconstructed image of the interaction which occurred between user and application, enriched by the duration of particular actions and responses.

Task no 4. Usability evaluation

Having measured and estimated all usability measures (see Section II), as well as the visualization of all relevant information, one can further analyze, classify and interpret the obtained outcome. On the other hand, some might go through the audio-video recordings, investigating these cause-and-effect relationships that may lead to a loss of effectiveness and a decrease in satisfaction.

In our opinion, the enriched action-response model provides an effective approach to evaluating particular usability attributes. At the interpretive level of research, the results of the quantitative analysis, provide explanations for evaluators to compile consistent and proper judgments.

The outcome of this step is the report, which in general presents results and conclusions, as well as a list of recommendations with applicable participant's reviews. In particular, the report categorizes usability issues into three groups: (1) bugs and errors, (2) design and (3) performance. In this line of thinking, the addressed groups may respectively concern testers, designers and developers.

IV. THE RVDA TOOL

To develop the RVDA tool (Retrospective Video Data Analyzer) we used the Electron open source library. Developed by GitHub, and devoted for building cross-platform desktop applications with HTML, CSS, and JavaScript, combines Chromium and Node.js into a single runtime environment [62]. We also used npm to manage packages for JavaScript runtime environment. In general, the above are the major devkit components.

The architecture of the tool has been conceived to allow for an easy and scalable video content analysis, supporting all common video file formats available for playback, where one or more files can be simultaneously opened and analyzed in separate windows.

From a technical viewpoint, there are main modules and rendering modules. The former is responsible for operations on windows, user interface properties, and communication with the operating system. The latter includes tasks on the video content and all associated functions. By design, neither communication nor data exchange occur within modules.

The user interface (UI) is divided into three sections: (1) the video preview area, (2) menu panel, and (3) display manipulators and the timeline. The size of the sections can be easily modified, along with the position of the tool on the computer screen.

Fig. 1 The RVDA Graphical User Interface

There are three media controls available (Stop, Play, Slow) to enact, change or adjust the process of playing video content. These controls support human cognition during processing video data that, in turn, advances the analyzing and extraction of data.

There are four levels available, which are used to break down the tasks to the user's actions and the application's responses (events). Accordingly, the Start and Stop buttons are used to drop pins on the timeline which mark the beginning and the end of each event. Moreover, there are also Save, Reset, and Clear features available.

The Export feature allows timeline data to be saved to the external CSV file which contains the event: identifier, name, start time, end time, and the level number. In this case, data obtained from a series of experiments can be consolidated for further reporting and analysis.

V. Experimental Setup

The inception of the laboratory was in June 2018, and it is located in the WSB University in Gdansk. Both hardware and software apparatus were chosen to meet specified requirements regarding project scope and budget.

Each application testing session begins with a short introduction, including goals and research agenda. Afterwards the usability testing procedure is described and executed. Our experience shows that a single session lasts approximately 15 minutes.

Hardware apparatus

For real-time image and voice capture and recording we used the Genee Vision 150 document camera. The design and specification combine five major components: optics, camera, lighting system and motherboard with firmware, and software tool.

Fig. 2 The Genee Vision 150 document camera

The built-in camera resolution is 2 megapixel, resulting in image dimensions of 1920 by 1080. While resolution is an important specification, those described above proved to be adequate to extract all required data in the context of the study. Optical and digital zoom is 8x and 10x, and the camera can rotate vertically 180º and 180º horizontally. The device is connected via a USB port to the local computer, equipped with the software apparatus and the latest hardware drivers and software libraries.

Software apparatus

In total we used four software tools. Firstly, the VideoCap captures audio-video data, transmitted by the document camera, to the mp4 file format. Secondly, the input mp4 data is analyzed under the RVDA tool. Thirdly, in order to document and to actively collaborate with all interested parties on the recognized usability issues, on the external virtual server we have installed, configured and deployed self-hosted Git service, namely Gogs. Fourthly, for advanced statistical analysis we take advantage of commerce and open-source tools to make sense of quantitative data.

VI. Methodology Testing and Validation

The elaborated methodology was tested from January to May 2019, in a group of four simultaneous student projects. Each project concerned a different set of mobile applications. Usability testing included five adult participants, both males and females with relevant knowledge and skills, who were using their own smartphones during the session.

Moreover, the methodology underwent one proof of concept (PoC), despite the preliminary stage of its maturity. During the PoC meeting, the interested parties have underlined its practicality, as well as the possible benefits to both designers and developers. Here, it is worth highlighting just one voice of appraisal: "*Despite economic constraints, analysis of video has strong software quality validity, offering rich insight into both application and user behavior. Studying the interactions in a retrospective manner provides valuable lessons and affirmation for quality assurance practices*".

VII. Conclusions

We are aware that the methodology needs further consultation and discussion with the software industry, with the goal of reaching an optimal cost-benefit ratio. From the economic perspective, the methodology is time-consuming and labor-intensive. While mobile application development projects have relatively low budgets, there may be no allowance to reserve sufficient funds for this area.

However, in conclusion, while the methodology has been acknowledged to be application-agnostic, fully adaptive and replicative, we believe that in the near future some interested parties will decide to implement our methodology, or at least incorporate some of its components.

Acknowledgment

I would like to thank the Rector of the WSB University in Gdansk, Professor Jan Wiśniewski, for his financial support and motivation to continue research, and Professor Małgorzata Rozkwitalska, the Head of the Management Department, for valuable insights and helpful discussions.

References

[1] McGee, M., Rich, A., & Dumas, J. (2004). Understanding the usability construct: User-perceived usability. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 907–911.

[2] Sung, J., & Yun, Y. (2010). Toward a more robust usability concept with perceived enjoyment in the context of mobile multimedia service. International Journal of Human Computer Interaction, 1(2), 12–32.

[3] Koohang, A. (2004). Expanding the concept of usability. Informing Science: International Journal of an Emerging Transdiscipline, 7, 129–141.

[4] Jeffries, R., & Desurvire, H. (1992). Usability testing vs. heuristic evaluation: was there a contest? ACM SIGCHI Bulletin, 24(4), 39–41.

[5] Zhang, Z., Basili, V., & Shneiderman, B. (1999). Perspective-based usability inspection: An empirical validation of efficacy. Empirical Software Engineering, 4(1), 43–69.

[6] Goldberg, J. H., & Wichansky, A. M. (2003). Eye tracking in usability evaluation: A practitioner's guide. The Mind's Eye, 493–516.

[7] Norman, D. A., & Draper, S. W. (1986). User centered system design: New perspectives on human-computer interaction. CRC Press.

[8] Desmet, P. M., & Hekkert, P. (2007). Framework of product experience. International journal of design, 1(1), 57–66.

[9] Plattner, H., Meinel, C., & Leifer, L. (Eds.). (2010). Design thinking: understand–improve–apply. Springer Science & Business Media.

[10] Instone, K. (1997). Usability heuristics for the Web. Web Review, October.

[11] Nielsen, J. (2005). Ten usability heuristics.

[12] Paz, F., Paz, F. A., Pow-Sang, J. A., & Collantes, L. (2014). Usability heuristics for transactional web sites. 11th International Conference on Information Technology: New Generations, 627–628. IEEE.

[13] Weichbroth, P. (2018). Delivering Usability in IT Products: Empirical Lessons from the Field. International Journal of Software Engineering and Knowledge Engineering, 28(07), 1027–1045.

[14] Manchester, J. (2013). Software Revolution, Part IV: Computing in a Mobile World. https://www.forbes.com/sites/truebridge/2013/10/22/software-revolution-part-iv-computing-in-a-mobile-world/#325625c71cd5 [2019-05-14].

[15] Truong, A. (2016). China saw a massive shift from desktop to mobile computing in 2015. https://qz.com/622543/china-saw-a-massive-shift-from-desktop-to-mobile-computing-in-2015/ [2019-05-14].

[16] Varshney, U. (2003). Issues, requirements and support for location-intensive mobile commerce applications. International Journal of Mobile Communications, 1(3), 247–263.

[17] Wasserman, T. (2010). Software engineering issues for mobile application development. FoSER 2010.

[18] Parsons, D., Ryu, H., & Cranshaw, M. (2006, July). A study of design requirements for mobile learning environments. In Sixth IEEE International Conference on Advanced Learning Technologies (ICALT'06), 96–100. IEEE.

[19] Nilsson, E. G. (2009). Design patterns for user interface for mobile applications. Advances in engineering software, 40(12), 1318–1328.

[20] Hoebeke, J., Moerman, I., Dhoedt, B., & Demeester, P. (2004). An overview of mobile ad hoc networks: applications and challenges. Journal-Communications Network, 3(3), 60–66.

[21] Anagnostopoulos, C. B., Tsounis, A., & Hadjiefthymiades, S. (2007). Context awareness in mobile computing environments. Wireless Personal Communications, 42(3), 445–464.

[22] Raptis, D., Tselios, N., Kjeldskov, J., & Skov, M. B. (2013, August). Does size matter?: investigating the impact of mobile phone screen size on users' perceived usability, effectiveness and efficiency. In Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services, 127–136. ACM.

[23] Nkosi, M., & Mekuria, F. (2011). Improving the capacity, reliability & life of mobile devices with Cloud Computing. In 2011 IST-Africa Conference Proceedings, 1–9. IEEE.

[24] Wajahat, N. (2019). Mobile App Performance Testing: CheckList, Tools (Andriod & iOS). https://www.guru99.com/mobile-app-performance-testing-strategy-tools.html [2019-05-17].

[25] Weichbroth, P. (2019). Usability of mobile applications: a systematic literature study (in print).

[26] ISO 9241-11:2018(en). Ergonomics of human-system interaction. Part 11: Usability: Definitions and concepts.

[27] Harrison, R., Flood, D., & Duce, D. (2013). Usability of mobile applications: literature review and rationale for a new usability model. Journal of Interaction Science, 1(1).

[28] Ghazizadeh, F. Z., & Vafadar, S. (2017). A quantitative evaluation of usability in mobile applications: An empirical study. International Symposium on Computer Science and Software Engineering Conference (CSSE), 1–6. IEEE.

[29] Kabir, M. A., Salem, O. A., & Rehman, M. U. (2017). Discovering knowledge from mobile application users for usability improvement: A fuzzy association rule mining approach. In 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), 126–129. IEEE.

[30] Krzewińska, J., Indyka-Piasecka, A., Kopel, M., Kukla, E., Telec, Z., & Trawiński, B. (2018). Usability Testing of a Responsive Web System for a School for Disabled Children. In Asian Conference on Intelligent Information and Database Systems, 705–716. Springer.

[31] Widyanti, A., & Ainizzamani, S. A. Q. (2017). Usability evaluation of online transportation' user interface. International Conference on Information Technology Systems and Innovation, 82–86. IEEE.
[32] Saleh, A., Ismail, R., & Fabil, N. (2017). Evaluating usability for mobile application: a MAUEM approach. Proceedings of the 2017 International Conference on Software and e-Business, 71–77. ACM.
[33] McDonald, K., Abell, W., Smith, C., & Gibbs, S. (2016). Lessons learned from evaluating a mobile app out in the field. 4th International Conference on User Science and Engineering (i-USEr), 5–10. IEEE.
[34] do Nascimento Mendes, I., & Dias-Neto, A. C. (2016). A Process-Based Approach to Test Usability of Multi-platform Mobile Applications. International Conference of Design, User Experience, and Usability, 456–468. Springer.
[35] Aljaber, T., & Gordon, N. (2017). A guidance and evaluation approach for mHealth education applications. International Conf. On Learning and Collaboration Technologies, 330–340. Springer.
[36] Parsazadeh, N., Ali, R., & Rezaei, M. (2018). A framework for cooperative and interactive mobile learning to improve online information evaluation skills. Computers & Education, 120, 75–89.
[37] Lyons, K. M. (2014). U.S. Patent Application No. 13/995,514.
[38] Sweller, J., & Chandler, P. (1991). Evidence for cognitive load theory. Cognition and instruction, 8(4), 351–362.
[39] Chynał, P., Szymański, J. M., & Sobecki, J. (2012). Using eyetracking in a mobile applications usability testing. Asian Conf. On Intelligent Information and Database Systems, 178–186. Springer.
[40] Ryan, C., & Gonsalves, A. (2005). The effect of context and application type on mobile usability: an empirical study. Proceedings of the Twenty-eighth Australasian Conf. 38, 115–124.
[41] Wang, Y., & Brennan, R. (2017). Building High Usability Consumer-oriented Linked Data Mobile Apps. AICS, 295–307.
[42] Korhan, O., & Ersoy, M. (2016). Usability and functionality factors of the social network site application users from the perspective of uses and gratification theory. Quality & Quantity, 50(4), 1799–1816.
[43] Biel, B., Grill, T., & Gruhn, V. (2010). Exploring the benefits of the combination of a software architecture analysis and a usability evaluation of a mobile application. Journal of Systems and Software, 83(11), 2031–2044.
[44] de Paula, D. F., Menezes, B. H., & Araújo, C. C. (2014). Building a quality mobile application: A user-centered study focusing on design thinking, user experience and usability. In International Conference of Design, User Experience, and Usability, 313–322. Springer.
[45] Tomaschko, M., & Hohenwarter, M. (2018). Usability Evaluation of a Mobile Graphing Calculator Application Using Eye Tracking. In International Conference on Learning and Collaboration Technologies, 180–190. Springer.
[46] Merriam-Webster English Dictionary, https://www.merriam-webster.com/dictionary/methodology [2019-04-06]
[47] https://www.merriam-webster.com/dictionary/method
[48] https://www.merriam-webster.com/dictionary/rule
[49] https://www.merriam-webster.com/dictionary/postulate
[50] Denzin, N. K. (2017). The research act: A theoretical introduction to sociological methods. Routledge.
[51] Thurmond, V. A. (2001). The point of triangulation. Journal of nursing scholarship, 33(3), 253–258.
[52] Brown, P. A. (2008). A review of the literature on case study research. Canadian Journal for New Scholars in Education.
[53] Sale, J. E., Lohfeld, L. H., & Brazil, K. (2002). Revisiting the quantitative-qualitative debate: Implications for mixed-methods research. Quality and quantity, 36(1), 43–53.
[54] Altheide, D. L. & Johnson, J. M. (1994). Criteria for assessing interpretive validity in qualitative research. In: Denzin, N. K. & Lincoln, Y. S. (Eds), Handbook of Qualitative Research. Thousand Oaks, CA: Sage Publications, 485–499.
[55] Holtz, P., & Odağ, Ö. (2018). Popper was not a Positivist: Why Critical Rationalism Could be an Epistemology for Qualitative as well as Quantitative Social Scientific Research. Qualitative Research in Psychology, 1–24.
[56] Mack, N. (2005). Qualitative research methods: A data collector's field guide.
[57] Miller, R. L., & Brewer, J. D. (Eds.). (2003). The A–Z of social research: a dictionary of key social science research concepts. Sage.
[58] Kasper, G. (1998). Analysing verbal protocols. Tesol Quarterly, 32(2), 358–362.
[59] Guterman, O., & Neuman, A. (2017). Different reasons for one significant choice: Factors influencing homeschooling choice in Israel. International Review of Education, 63(3), 303–318.
[60] Lockton, D., Harrison, D., Holley, T., & Stanton, N. A. (2009). Influencing interaction: development of the design with intent method. 4th International Conf. on Persuasive Technology, ACM.
[61] Van Den Haak, M., De Jong, M., & Jan Schellens, P. (2003). Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue. Behaviour & information technology, 22(5), 339–351.
[62] Electron Documentation, https://electronjs.org/docs/tutorial/about.

14th Conference on Information Systems Management

THIS 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
- **Ziemba, 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
- **Eisenbardt, Monika,** University of Economics in Katowice, Poland, Poland
- **El Emary, Ibrahim,** King Abdulaziz Univetrsity, Saudi Arabia
- **Espinosa, Susana de Juana,** University of Alicante, Spain
- **Fantinato, Marcelo,** University of Sao Paulo, Brazil
- **Gabryelczyk, Renata,** University of Warsaw, Poland
- **Gaweł, 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

Communication Papers of the Federated Conference on Computer Science and Information Systems pp. 109–117
DOI: 10.15439/2019F63
ISSN 2300-5963 ACSIS, Vol. 20

Improved Virus Optimization algorithm for two-objective tasks Scheduling in Cloud Environment

Kadda Beghdad Bey[1], Sofiane Bouznad[1], Farid Benhammadi[1] and Hassina Nacer[2]

[1]*Laboratory of Distributed and Complexes Systems, Ecole Militaire Polytechnique, BP 17 Bordj el Bahri, Algiers, Algeria*
(k.beghdadbey, bouznad.sofiane and fbenhammadi2008)@gmail.com

[2]*MOVEP Laboratory. Computer Science Department University of Science and Technology, USTHB, Algiers, Algeria*
sino nacer@yahoo.fr

***Abstract* — *Cloud computing* is increasingly recognized as a new way to use on-demand, computing, storage and network services in a transparent and efficient way. The development of applications in cloud environments is faced with the need to efficiently schedule a large number of tasks and resources. However, in the most of the time, the resources in cloud are not efficiently utilized due to inadequate scheduling task algorithm in virtual machines. Therefore, task scheduling is one of the most challenging issues in cloud computing. In this paper, we propose two-objective virus optimization algorithm of the makespan and the cost, for mapping tasks to virtual machines in order to meet the needs of cloud service quality and proper assignment of resources. Thus, based on genetic algorithm some parameters of Virus optimization algorithm are redefined to strengthen sorting ability between virus infection strategies. Our combined methods aims to improve the performance of scheduling algorithms. It outperforms some existing approaches for task scheduling in Cloud computing.**

Keywords: *Independent task scheduling, Cloud computing, Virus optimization algorithm, Genetic algorithm, Two-objective optimization, CloudSim simulator.*

I. INTRODUCTION

Within few years, IT industries start using Cloud Computing (CC) by serving on demand requests of the users with self-managed virtual infrastructure and with efficient resources utilization. Growth of cloud computing slower down the efficiency, throughput and resource utilization for which cloud computing need to be evolve. Task scheduling is considered as one way to enhance the efficient resource utilization in cloud environments. In fact, independent computational tasks are supposed to be executing in parallel when they are executed concurrently on different virtual machines. The scheduling strategy defines the instants when the algorithm is called to produce a schedule based on the resources performances forecasting and independent computational tasks to be executed. The task scheduling strategies can be classified on two different types: static and dynamic. Static strategies define a schedule at compile or at launch time based on the knowledge of the processors availability and tasks to execute. Dynamic strategies, are applied when the tasks arrival time is not known a priori and therefore the system needs to schedule tasks as they arrive [1].

Plethora of multi-objective optimization techniques for independent tasks scheduling has been proposed for assignment of tasks to machine in Cloud Computing systems [2-11]. The Min-Min heuristic is the well-known standard scheduling algorithm for its performance, simplicity and practicability. The Min-Min heuristic [3] gives the highest priority for dynamic tasks scheduling in heterogeneous computing systems. The principle of Min-Min heuristic is to finish each task as early as possible and it schedules the tasks with the selection criterion of minimum earliest completion time.

In this paper, we are interested in two-objective optimization for scheduling independent tasks in cloud computing environment. The optimization process consists to minimize the makespan value and the operational cost in order to ensure the performance and quality of service in the cloud. The proposed optimization approach uses a virus optimization algorithm [10]. Thus to provide an improvement of this algorithm for task scheduling purpose, we propose GA operators to select strong and weak virus for exploitation and exploration in space search. In addition, the initialization step is based on standard scheduling algorithms such as Min-min, Max-min and Tabou. Experimental results show that our proposed approach improves the performance of several existing approaches in literature such as the Min-Min (Max-Min) heuristic, Genetic heuristic and Virus optimization algorithms. Moreover, we present in this paper a comparative study in which the proposed scheduling approach is evaluated with improved Particle Swarm Optimization (PSO) approach [12] using a set of different types of Expected Time to Compute (ETC) matrices up to randomly heterogeneous machines and heterogeneous tasks using simulated and real data.

The remainder of this paper is organized as follows. Section 2 describes the relevant multi-objective optimization scheduling problems in Cloud environment. Section 3 formulates the two-objective optimization problem for independent tasks scheduling. Section 4 details the proposed standard combination, Virus optimization and genetic algorithms used for resolving this problem. Experimental results and performance evaluations of the proposed combination are reported in section 5. Finally, section 6 presents conclusions and directions for future works.

©2019, PTI

II. RELATED WORKS

Various classical scheduling tasks algorithms have been proposed and deployed until date in CC environments, such as: First Come First Serve, Min-Min, Min-Max based scheduling. These classical schemes posed drawbacks of resulting into more execution time and reduced throughput. Recently, several researches works focus on multi-objective optimization tasks scheduling methods in CC and proposes several techniques. Most of them are based on evolutionary algorithms to minimize the operational cost and ensure the performance and quality of service in CC. A vast literature exists on bio inspired approaches for optimized scheduling tasks in CC. Ahmad et al. [11] proposed a hybrid genetic algorithm for solving the workflow-scheduling problem and optimizing the load balance for maximum resource utilization. The Multi-objective particle swarm optimization is another class of task scheduling problem that has been addressed in CC environment [12, 13]. Generally, the algorithms based on this optimization achieved best performances compared to the classical scheduling methods. In recent times, others methods use the combination of several evolutionary strategies. This combination of evolutionary type optimization algorithms has provided the best ways to solve multi-objective optimization problems, because of their efficiency, robustness and quick convergence compared to strategies using only single evolutionary optimization method. In 2015, Shu, and al. [14] proposed an improved clonal selection algorithm for meeting the service level agreement requested by the users. The experimental results show that the proposed algorithm performs better than other two algorithms with minimum execution time and increased throughput of the cloud computing system. In [15], authors proposed an algorithm based on the combination of genetic algorithm along with fuzzy optimization theory. Another multi-objective optimized tasks scheduling algorithm using genetic algorithms with greedy approach is proposed in [16]. This algorithm not only performs task scheduling but also perform others load balancing methods in Cloud environment. Chu [17] used the combination of genetic algorithm and support vector machines for two-objective optimized tasks scheduling algorithm. Based on the cross operation of genetic algorithm and components selection of partial regression, the proposed work gives a high effective scheduling and service cost may be reduced in cloud computing environment using both completion time and cost of tasks. Similarly, Kim et al. [18] have developed biogeography-based optimization for tasks scheduling. This algorithm performs more satisfactorily than other optimization algorithms, such as genetic algorithm and particle swarm optimization, in large size problems. Lakra et al. [19] proposed a Two-objective tasks scheduling algorithm for mapping tasks to a VMs improving the throughput of the datacenter and reducing the cost without compromising the service level agreement in cloud environment. The proposed scheduling algorithm in [20] is involving Non-dominated sorting which targets the two-objective issues considering completion time and cost minimizations. The exposed results by the authors outperform the preceding algorithms and this method represents better convergence performance and resource scheduling capability in the different number of resources. These combination strategies perform satisfactorily than other multi-objective optimization problems using only genetic algorithm or particle swarm optimization.

III. PROBLEM DEFINITION

In cloud computing, many datacenters consist of several servers where each server runs a number of virtual machines that have different capacity to execute tasks with different QoS parameter. Scheduling refers to the mapping or assigning a task to a specific virtual machine, such that resource utilization increases.The main problem is to bind set of tasks received by the broker to the virtual machines with the respect of optimized QoS. Based on the makespan and the cost, this problem can be modelled using two-objective optimization for tasks scheduling problem. This optimization model includes two objective functions to be minimized simultaneously:

$$Find\ X = \{x_1, x_2, \ldots\ldots\ldots x_n\}$$

Which minimize two objective functions f_1, f_2 *:Min*$(f_1(x), f_2(x))$, where X is the feasible solutions set.

Our two-objective model for task scheduling optimization in CC environment can be described by a triplet (*T, VM, F*). $T = \{t_1, t_2, \ldots, t_n\}$ is a consumer tasks set including *n tasks*, $VM = \{vm_1, vm_2, \ldots, vm_l\}$ is a virtual machine resource set including l virtual machine and *F={Makespan, Cost}* is a set of the considered cloud resource scheduling and optimization functions. The proposed model is defined as follows:

The *Makespan* is the maximum completion time of all tasks in all virtual machines. It can be obtained using the Expected Time to Compute (*ECT*) matrix by the following equation:

$$Makespan = \underset{1 \le j \le m}{Max} \sum_{i=1}^{n} \left(ETC_{ij} \times x_{ij} \right)$$

Where :

$$x_{ij} = \begin{cases} 1 & \text{if a task } t_i \text{ will run on } vm_j \\ 0 & \text{otherwise} \end{cases}$$

- ETC_{ij} is the estimated execution time of task t_j on machine vm_i and the values of ETC matrix are calculated by the following equation:

$$ETC_{ij} = \frac{L_i}{mips_j}$$

L_i indicates the number of instructions required by the task T_i (implementation time) and $Mips_j$ indicates the frequency of cloud computing virtual machine vm_j.

However, the *Cost function* in our tasks scheduling problem is based on the cost C_j of each virtual machine vm_j. The total execution cost for all tasks scheduling can be defined according to the following formula:

$$C = \sum_{j=1}^{l} \sum_{i=1}^{n} ETC_{ij} \times x_{ij} \times C_j$$

Above all, we need to find the most reasonable tasks scheduling which minimizes the makespan and the cost in cloud system. Thus, the two-objective tasks scheduling problem can be modeled by the aggregation (Weighted Sum Method) of objective optimization functions defined as:

$$Z(x) = Min(\lambda \times Makespan + (1-\lambda) \times C)$$

under the constraints:

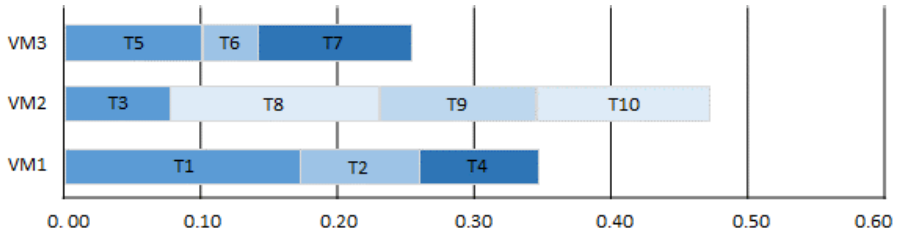
$$\begin{cases} \sum_{j=1}^{m} x_{ij} = 1 \quad \forall i \in T \\ x_{ij} \in \{0,1\} \\ 0 \leq \lambda \leq 1 \end{cases}$$

Where λ and (1-λ) represents the weights of the makespan and the cost in our two-objective function.

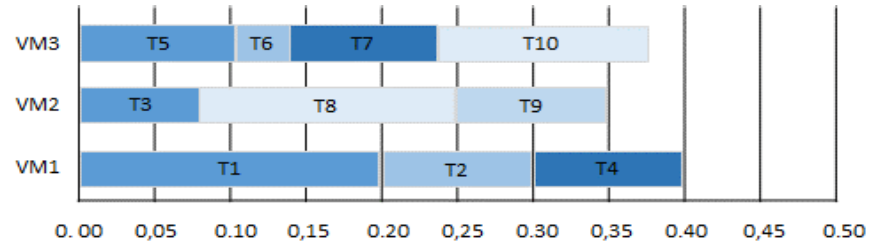
In two-objective optimization, a sufficient solution that minimizes these two objective functions at the same time can considers their linear combination (aggregation called Weighted Sum Method) or Pareto optimal solutions. These solutions cannot be enhanced for any objective without degrading minimum of the other objective. For example, consider the following case where a scenario of *ETC* matrix defined by ten tasks and three virtual machines as given in Table 1. According to the two task assignments (Figure 1), we obtain two solutions (Makespan=0.68, Cost=80.13$) and (makespan=0.40, Cost=82.8$). Therefore, these solutions show that the makespan objective can degrade the cost objective in the solution selection. Thus, we can select one solution based on the λ value according to the weights of these two objectives.

TABLE I. AN EXAMPLE OF ETC VALUES

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	*Cost($)*	*Mips*
Vm_1	0.10	0.20	0.05	0.10	0.15	0.06	0.15	0.10	0.06	0.20	60	1000
Vm_2	0.17	0.33	0.08	0.17	0.25	0.10	0.25	0.17	0.10	0.33	40	600
Vm_3	0.07	0.13	0.03	0.07	0.10	0.04	0.10	0.07	0.04	0.13	120	1500



(a) Scenario 1



(b) Scenario 2

Fig 1. Two scenarios of tasks scheduling in CC.

IV. PROPOSED META-HEURISTIC APPROACH: MOVOA

The development of applications for CC environments is being challenged by the need of scheduling a large number of tasks, datasets and resources efficiently [21]. The general problem of optimally mapping tasks to machines has been shown to be NP-complete. In this work, we propose a combined evolutionary heuristics for solving independent task scheduling problems in CC environment. This combination uses the two-objective optimization methods based on adapted virus optimization algorithm and genetic operators. This combination allows finding the task assignment that minimizes the makespan and the cost. In this section, we briefly give a description of the original virus optimization algorithm [10] and thereafter we introduced our tasks scheduling method based on improvement of this algorithm.

A. Virus Optimization Algorithm

The Virus Optimization algorithm is a meta-heuristic technique, population-based solution used to solve many optimization problems with single objective like continuous domain problems [10]. This algorithm imitates the behavior of the viruses attacking a living cell by infection. Once they are entered, they will start replicating and alter the genetic material of the host cell. More viruses will be produced and ultimately host cell will die. In this algorithm, solution space is taken as cell itself and global optima can be found inside the cell. Many viruses can coexist within a host cell and each such virus represents a solution in the solution space. The viruses are divided into two categories: Strong and Weak, which corresponds to the exploration and exploitation capability of the virus optimization algorithm. Strong viruses will have high objective function value compared to weak viruses and they will replicate faster than them. The algorithm does mainly three types of phases: initialization, replication, updating and maintenance. An advantage of this algorithm is that it can be easily parallelized and therefore easily implemented.

The initialization step occurs according to the fitness value and consist to create the starting population of the possible solutions (virus) fulfilling the constraints. These solutions are sorted based on the objective function evaluation to select strong and weak virus [10]. In the replication step, VOA generates new virus where are ranked in strong and weak ones according to some objective function values. For example, if the required number of strong viruses is three, then the first three virus of the created list are considered strong while the remaining are considered weak. In the last step, once the new viruses have been generated using the replication process, VOA checks and maintain the new population size using the corresponding objective function values. As a result, only the strong viruses are kept ordered according to their objective function values. The checking process verifies the convergence and VOA determines whether the exploitation has to be intensified by creating new members closer to the stronger viruses [10]. However, the Maintain process (antivirus) is activated by interaction between the viruses and the host cell; the antivirus is triggered at each replication, killing a given number of viruses according to some fixed parameters such that the number of strong virus, weak virus and population size [10]. For example, if the needed number of strong viruses is three, then the first three viruses in the created virus list are considered strong while the remaining viruses are considered weak. In addition, this process eliminates some virus from the population if the total number of viruses inside the host cell exceeds 1000 virus. Finally, the stopping condition of the VOA algorithm can be the maximum number of the iterations or the maximum number of the replications.

B. Task scheduling based on improved VOA

Based on the original VOA, we adapted modifications in order to enhance the performances of initialization and replication steps. This following section presents the description to accomplish to the VOA adaptation for two-objective optimization based tasks scheduling in CC, called MOVAO. The population initialization step uses standard algorithms such as Min-Min, Max-Min and Tabou to generate the initial population. In the modification step, a combination of the genetic operators in the virus replications of the classical VOA is applied. Fig.2 shows the visual representation for the main steps of the proposed two-objective optimization process based on VOA and genetic algorithm.

The presented diagram consists of four steps: Initialization, classification, replication and antivirus (maintenance). The proposed two-objective tasks scheduling algorithm is described in Algorithm 1. First, we defined the control parameters of the proposed MOVOA as follows:

- $\boldsymbol{Pop_{size}}$: *Size of the initial population;*
- $\boldsymbol{Pop_{virus}}$: *Maximum number of viruses in a cell;*
- $\boldsymbol{N_{task}}$: *Number of tasks to be performed;*
- $\boldsymbol{N_{vms}}$: *Number of virtual machines;*
- $\boldsymbol{P_{virusStrong}}$: *The proportion of strong viruses in a population;*
- $\boldsymbol{N_{virusStrong}}$: *Number of strong viruses;*
- $\boldsymbol{N_{virus}}$: *Number of viruses to be removed in the antivirus process;*
- $\boldsymbol{N_{repStrong}}$: *Number of viruses to generate from a strong virus;*
- $\boldsymbol{N_{repWeak}}$: *Number of viruses to generate from a weak virus;*
- $\boldsymbol{ETC}$: *Matrix of task execution times in virtual machines;*
- $\boldsymbol{C}$: *Vector of execution cost on each virtual machine;*
- $\boldsymbol{S_{best}}$: *Vector representing the best solution found.*

In our proposed MOVOA, the population initialization is based on some standard scheduling algorithms such as Min-Min, Max-Min and Tabou. According to the logic of the VOA, this first step is to create the starting population of the possible solutions fulfilling the constraints of these standards algorithms.

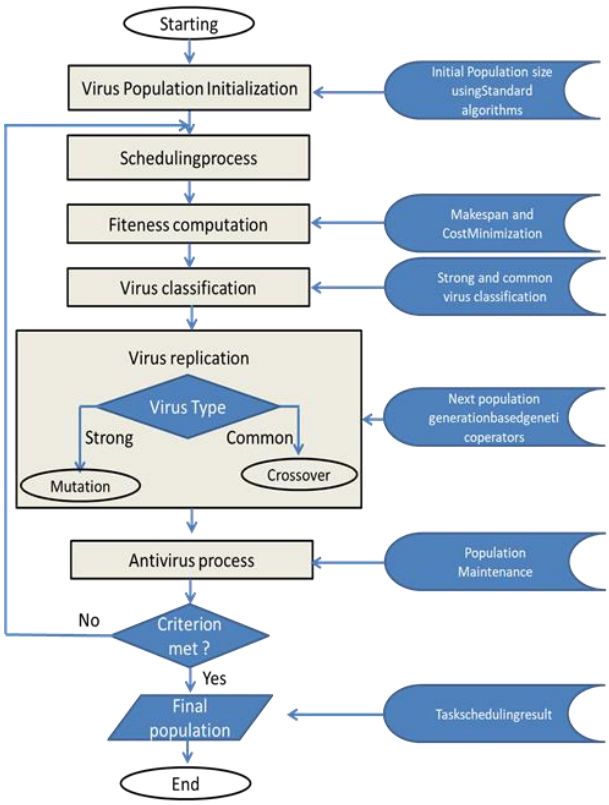


Fig.2 MOVOA Flow chart for task scheduling in CC.

Algorithm 1: MOVOA Algorithm

1 : **Input :** Pop_{size}, N_{tasks}, N_{VM}, $P_{VirusStrong}$, ETC, C, Pop_{Virale};
2 : **Output :** S_{best};
3 : $Pop[Pop_{size}]$ = **Initialze_Population**(Pop_{size}, N_{tasks}, N_{vm}) ;
4 : Moy = **Evaluate_Fitness**(Pop, ETC) ;
5 : i = 1 // Iterations number;
6 : **while** (i <= Iteration-max) **do**
7 : **classification**(Pop, $P_{VirusStrong}$) ;
8 : [$N_{rep\text{-}strong}$, $N_{rep\text{-}low}$] = random(1,10); // Avec $N_{rep\text{-}strong}$>$N_{rep\text{-}low}$
9 : Pop_{new}= **replication** (Pop, $N_{rep\text{-}strong}$, $N_{rep\text{-}low}$) ;
10 : Calculate the viruses number to be eliminated according to the equation:
N_{virus}= *rand* (0, $Population_{\sout{size}} - Strong_{members}$);
11 : **antivirus** (Pop_{new}, N_{virus}) ;
12 : $Average_{new}$= **Evaluate_Fitness**(Pop_{new}, ETC) ;
13 :**if** ($Average_{new}$>$Average$) **then**
14 : Increment intensity by1;
15 : **endif**
16 : **if** (size(Pop_{new}) >Pop_{virale}) **then**
17 : **reduction** (Po_{pnew}) ;
18 : **endif**
19 : Pop = Pop_{new};
20 : $Average$ = $Average_{new}$;
21 : i = i + 1 ;
22 : **endwhile**
23 : return the best solution of Pop : S_{best};

Once the starting population has been initialized, each virus of the population will must be classified as strong or weak virus according to their objective function values and some predefined parameters (number of strong viruses). As shown in algorithm 2, this classification process is based on the fitness function (two-objective function $Z(x)$) and the strong virus number. For example, if we fix the strong viruses number to three, so the first three viruses in the created list are then considered strong while the remaining viruses are considered weak.

Algorithm 2: Population classification into strong and weak viruses

1 : **Input :**Pop, $P_{virusStrong}$;
2 : Trier(Pop) according to the fitness function;
3 : $N_{virusStrong}$= size(Pop) ×$P_{virusStrong}$;
4 : **for** i = 0 **to** size(Pop) **do**
5 : **if** (i <$N_{virusStrong}$) **then**
6 : setType(Pop[i], strong) ;
7 : **else**
8 : setType(Pop[i], weak) ;
9 : **endif**
10 : **endfor**

After classifying the viruses, each ranked virus in the population will reproduce by creating new viruses. Thus, two strategies for this reproduction are adopted. For the strong virus, four viruses and only three new viruses for weak are generated. The reproduction process based on improved replication function is designated in Algorithm 3.

The improvement uses two genetic operators according to virus types. In order to avoid the local convergence for strong virus reproduction (local solution), the mutation operator is used for these types of viruses. However, one-point and two-point crossover operators are used only for weak viruses. For example, for the two-point crossover operator, we generate two new viruses from two weak viruses, which are randomly selected from the population, as illustrated in figure Fig.3.

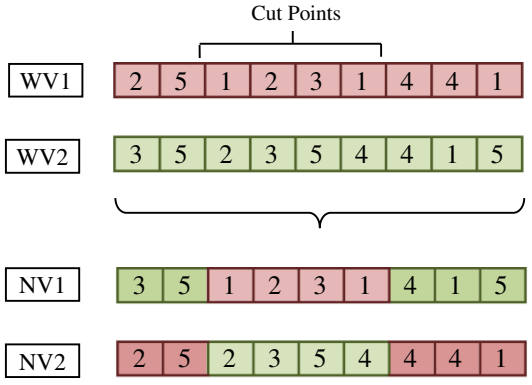


Fig.3 An example for crossover operation for two weak viruses

Algorithm 3: Improved replication with genetic operators

```
1 : Input :Pop, N_repStrong, N_repWeak;
2 : Output :Pop_new;
3 : Pop_new= Pop ;
4 : for i = 0 to size(Pop) do
5 : if (Type(Pop[i]) == strong)then
6 : Insert (listVirusStrong, Pop[i]) ;
7 : else
8 : Insert (listVirusWeak, Pop[i]) ;
9 : endif
10 : endfor
11 : while ( listVirusWeak! = Empty ) do
12 : for i = 0 to N_repWeak do
13 : // Remove two viruses from the list of weak viruses
14 : [virus1, virus2] =Insert (listVirusWeak) ;
15 : [fils1, fils2] =Crossing (virus1, virus2) ;
16 : Insert (Pop_new, [fils1, fils2]) ;
17 : endfor
18 : endwhile
19 : while ( listVirusForts! = V ide ) do
20 : for i = 0 to N_repStrong do
21 : // Remove one viruses from the list of strong viruses
22 : virus1 = Remove (listVirusStrong) ;
23 : fils1 = Mutation(virus1) ;
24 : Insert (Pop_new, fils1) ;
25 : endfor
26 : endwhile
```

After the replication process, the virus maintenance process execution (Antivirus algorithm) killed a given number of viruses according to the original mechanism [10] among the weak viruses set, because the weak viruses performing worse. In addition, if the total number of viruses inside the host cell exceeds 1000, generally the maintenance process will reduce the population size to the amount set initially defined (1000 viruses in our case). This extermination is based on fitness function values as a criterion performance of the virus population. The proposed antivirus process is described in Algorithm 4. Finally, if population performances did not improve after some replication process, the algorithm stops.

Algorithm 4: Principe of antivirus function ().

```
1 : Input :Pop, N_virus ;
2 : Output :Pop_new;
3 : Pop_new = Pop ;
4 : Trier(Pop_new) according to the fitness function;
5 : for i = 0 to N_virus do
6 : Eliminate the last virus of Pop_new ;
7 : endfor
```

V. RESULT AND EXPERIMENTS

In this section, an evaluation and comparison to validate the proposed strategy for task scheduling based MOVOA in Cloud Computing is presented. The conducted experiments use an Intel®Core™ i7-2600 machine. Moreover, we ran a set of experiments to compare it with the other heuristic-based approaches for independent task scheduling using the Cloudsim. The proposed approach has been applied to simulated and real data, with 12 different types of ETC matrix up to 16, 32, 64, 128 and 256 heterogeneous machines, and up to 512, 1024, 2048, 4096 and 8182 randomly generated heterogeneous tasks used in [2]. Thus, we evaluate the proposed MOVOA algorithm in order to minimize two objectives that are the makespan and the cost. These different types of ETC matrix are generated based on the following properties [2]:

Task heterogeneity – represents the amount of variance among the execution times of tasks for a given machine. The task heterogeneity is defined as: lo: low and hi: high.

Machine heterogeneity – represents the variation among the execution times for a given task across all the machines. The machine heterogeneity is defined as: *lo: low and hi: high.*

Consistency – an ETC matrix is said to be consistent whenever a machine m executes all tasks faster than another machine and the inconsistency if the machine m may be faster than another machine for some tasks and slower for others. The consistency type is defined as: *c: consistency; s: semi-consistency and i: inconsistency.*

For real data, we have created many VMs and tasks with different task size using log file introduced in [24]. Task size ranges from 100 to 10000 and the virtual machines from 3 to 48. The VMs have been created with the same processing power MIPS range.

First, we give the results of the improved initialization process adopted in our MOVOA. For simulated data, the obtained results in Table 2 indicate that the standard algorithms (Min-max, Min-Min and Tabou) in initialization process outperform the original VOA. Thus, it is important to note that our improved initialization process in VOA permits to give this algorithm competitive with standard algorithms presented in the literature.

For the setting of the three MOVOA parameters such as growth rate of weak viruses (B) and growth rate of strong viruses (B) and strong virus rate in the population (C), Pareto-optimal front is adopted, where each parameter and their combination are tested. Fig.4 illustrates the influence results of these parameters. As can be seen, the growth rate of strong viruses is the more important parameter for our MOVOA.

Fig.5 shows the influence of the strong virus rate on MOVOA convergence. Experimental results from this figure show that the strong virus rate in the population (C) has better convergence speed and search ability in solving the task scheduling problem. Our simulation parameters setting give the final results adopted for our MOVOA approach. Table 4 summarizes the values of these parameters.

TABLE II. RESULTS OF IMPROVED VOA IN INITIALIZATION PROCESS

Benchmarks		VOA	VOA-MaxMin	VOA-MinMin	VOA-TS
512×16	A.u_c_hihi	33515841,2075974	**9478748,54705373**	11641581,3124835	40734849,2597272
	A.u_c_hilo	3737479,59860668	**139640,463361260**	890323,882382025	717846,168886722
	A.u_c_lohi	9895642,40071556	**325969,379074195**	3548603,69777885	1806764,43897363
	A.u_c_lolo	118682,944348105	**4741,88496593059**	24576,3120032209	186154,388470771
	A.u_i_hihi	54581417,6358823	**5744281,35952104**	46888394,9179914	54177713,8001841
1024×32	A.u_c_hihi	4383867358,88497	**81168110,2585890**	597798515,574730	6242746503,06729
	B.u_c_hihi	1246601703,33311	**24296362,6739020**	191214363,403808	1736816774,16309
	B.u_i_lolo	536609,894367258	**1825,32098972682**	39076,0630305110	553247,143843531
	B.u_s_hi	1620976031,0024	**14406243,8979040**	207133649,666086	1631866310,84382
	B.u_s_hilo	16045273,0005123	**161710,828037789**	2246707,62102429	16504173,9119531
2048×64	A.u_c_hihi	10540814353,3380	**68382087,7694320**	4066453996,05607	21023179983,7550
	B.u_c_hihi	1020053393,50667	**6809804,53715252**	429375587,707337	1961912055,42496
	B.u_s_hilo	34185834,4073316	**156508,279151177**	2833080,10334692	35777433,2609869
	B.u_i_lolo	1150676,97244939	**1062,26755656421**	40530,6137005020	1160879,21708258
	B.u_s_hihi	3326676485,82300	**14012103,8195210**	286662462,690487	3523788387,30358
4096×128	A.u_c_hihi	23996285871,9060	**6624387,60572100**	399220512,895295	24316863996,2840
	B.u_c_hihi	2411310541,92293	**559791,014140813**	57468211,0218529	2449266772,19776
	B.u_i_lolo	2411851,97259703	**553,207997194039**	41703,9413734879	2458020,21421099
	B.u_s_hilo	71071822,1568924	**132058,746687681**	4053296,88793036	73666932,4830240
	B.u_s_hihi	7163570672,51973	**12989954,2801580**	370826776,222161	7359095019,85810
8192×256	A.u_c_hihi	50062019882,6850	**2855306,64712900**	410602822,406195	50514086017,2900
	B.u_c_hihi	146893067,056463	**936711,747171347**	123165194,094830	14813093403,2000
	B.u_i_lohi	4946876,98939152	**336,242668651566**	41518,2627413454	5005558,14873186
	B.u_s_hilo	149517372,311974	**109082,878871015**	5591057,69139399	15091076305,0000
	B.u_s_hihi	26761574493,4720	**3612402,92343436**	118162914,600465	28141746047,1400

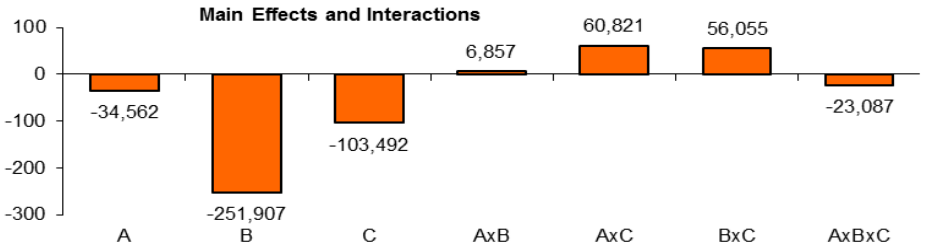


Fig.4 The setting of the three MOVOA parameters

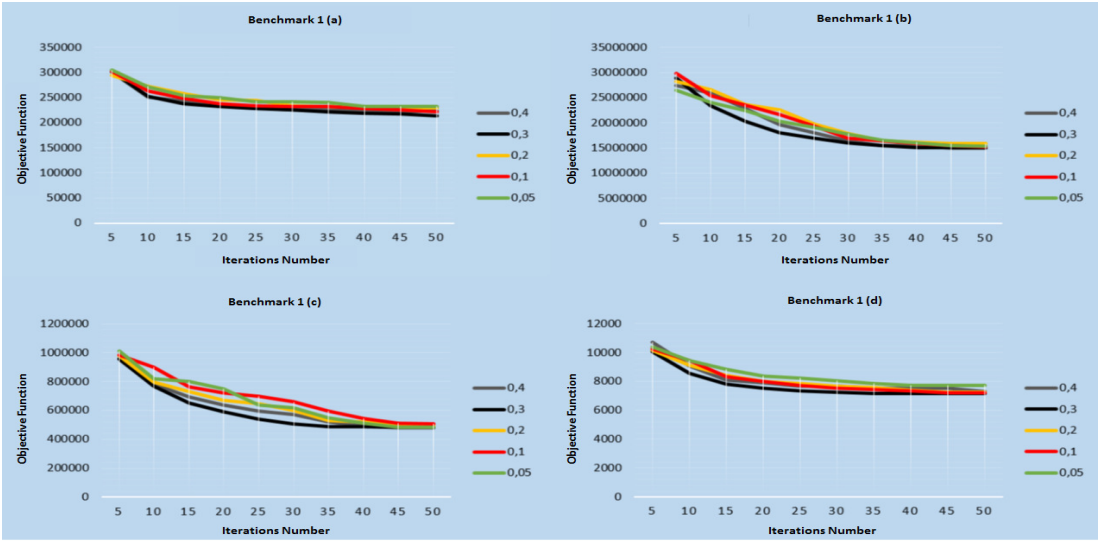


Fig.5 Influence of strong virus rate in the population.

Table 3 Parameter setting values

Parameters	Value
$P_{virusStrong}$	33.33% of classified population
Pop_{size}	50 virus
Pop_{virus}	Intul
$Iterations_{Max}$	30 Iterations

After tuning the parameters, we compare our MOVOA with some evolutionary algorithms with different values of the parameter λ and using a real instances. There are many prior works on multi-objective optimization problems for independent task scheduling using evolutionary techniques. The most popular of meta-heuristic algorithms are Genetic Algorithm (GA) and Particle Swarm Optimization (PSO). The comparison results with three competing algorithms (standard Max-Min algorithm and 2 evolutionary algorithms) are shown in Fig. 6. The results show that the MOVOA algorithm outperforms all the algorithms and still provides stable performance when the benchmark functions are subject to a variation of the two-objection linear combination parameter λ. Once again, MOVOA with Min-max initialization not only outperforms these evolutionary algorithms, but also shows outstanding results for all instances when λ=0.75 (Fig. 10(c)). Note that MOVOA performs the best in all of the thirteen instances except for 800×24.

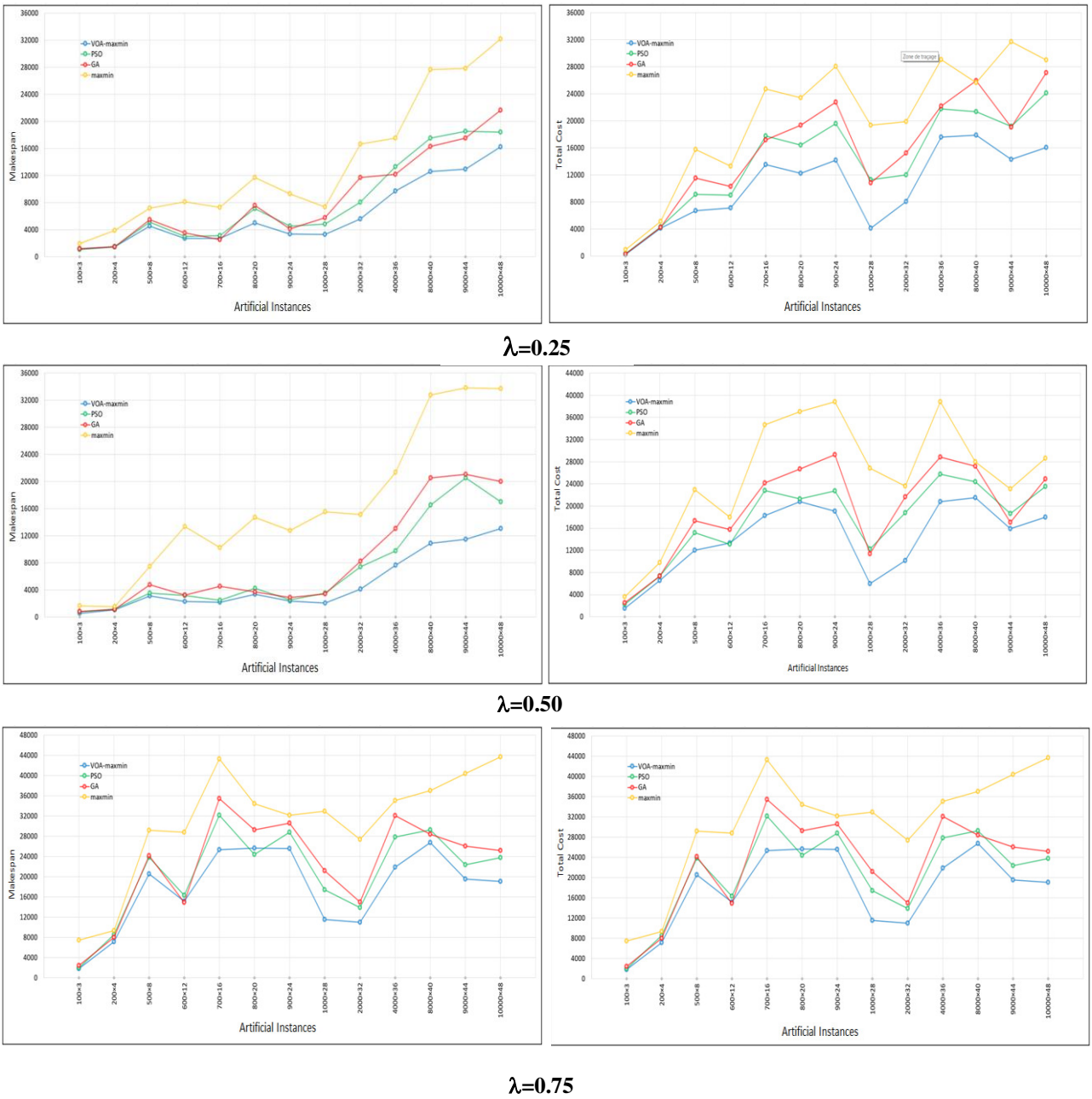


Fig. 6 Comparisons of MOVOA algorithm with Max-Min, GA and PSO algorithms.

VI. CONCLUSION

Scheduling of tasks is one of the most challenging problems in cloud computing environment. The proposed two-objective optimization for tasks scheduling algorithm in CC environment is based on the combination of virus optimization and genetic algorithms. This last algorithm gives a best replication for strong and weak virus in space search. Moreover, Max-Min algorithm reinforces the population initialization process for the proposed approach. The proposed algorithm has been simulated and the results are compared with standard and evolutionary algorithms previously implemented multi objective tasks scheduling algorithm for CC environment. Based on Weighted Sum Method and Pareto-optimal front, the proposed MOVOA showed competitive performance in terms of average two-objective function value, achieving the best results in most of the tested instances and outperforms some standard and evolutionary algorithms. The proposed algorithm can be generalized by taking consideration of some other QoS parameters of the CC environment.

REFERENCES

[1] J. Barbosa and B. Moreira, "Dynamic job scheduling on heterogeneous clusters", Eighth International Symposium on Parallel and Distributed Computing, 2009.

[2] O.H. Ibarra, C.E. Kim, Heuristic algorithms for scheduling independent tasks on non-identical processors, Journal of the ACM 24 (2) (1977), pp 280–289.

[3] H. Izakian, A. Abraham, V. Snasel, "Comparison of Heuristics for Scheduling Independent Tasks on Heterogeneous Distributed Environments", In: IWHGA '09: Proceedings of the IEEE International Workshop.

[4] S. Tareghian, Z. Bornaee, "Algorithm to improve job scheduling problem in cloud computing environment", 2nd International Conference on Knowledge-Based Engineering and Innovation (KBEI). IEEE, 2015, pp 684-688.

[5] C.Y. Liu, C.M. Zou and P. Wu, "A Task Scheduling Algorithm Based on Genetic Algorithm and Ant Colony Optimization in Cloud Computing", International Symposium on Distributed Computing and Applications To Business, Engineering and Science. 2014, pp 68-72.

[6] F. Tao, Y. Feng, L. Zhang, et al., "CLPS-GA: A case library and Pareto solution-based hybrid genetic algorithm for energy-aware cloud service scheduling". Journal of Applied Soft Computing, 2014, 19(6), pp 264–279.

[7] M. Zhang, Y. Yang, Z. Mi, et al., "An Improved Genetic-Based Approach to Task Scheduling in Inter-cloud Environment[, Ubiquitous Intelligence and Computing", IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATCScalCom), 2015: 997-1003.

[8] B. Keshanchi, A. Souri, N.J. Navimipour, "An improved genetic algorithm for task scheduling in the cloud environments using the priority queues: Formal verification, simulation, and statistical testing", Journal of Systems and Software, 2017, 124, pp 1-21.

[9] K. Beghdad Bey, F. Benhammadi, M. Y Boudaren and S. Khamadja, "Load Balancing Heuristic for Tasks Scheduling in Cloud Environment", 19th International Conference on Enterprise Information Systems (ICEIS2017), 26-29 April 2017.

[10] Y.C. Liang and J. R. C. Juarez, "A novel metaheuristic for continuous optimization problems: Virus optimization algorithm". In: Engineering Optimization 48.1 (2016), pp.73–93.

[11] S. G. Ahmad, C. S. Liew, E. U. Munir, T. F. Ang, S. U. Khan, "A Hybrid Genetic Algorithm for Optimization of Scheduling Workflow", Applications in Heterogeneous Computing Systems, Vol. 87, January 2016, pp. 80-90.

[12] L. Guo, G. Shao, and S. Zhao, "Multi-objective Task Assignment in cloud computing by Particle Swarm Optimization". In Proceedings of 8th Int. Conf. on Wireless Communications, Networking and Mobile computing, 2012, pp 1-4.

[13] S. Pandey, L. Wu, S. Guru and R. Buyya, "A particle swarm optimization based heuristic for scheduling workflow applications in cloud computing environments". 24th IEEE Int'l Conference on Advanced Information Networking and Applications (AINA), Perth, Australia, 2010, pp. 400-407.

[14] W. Shu, W. Wang, Y. Wang, "A Novel Energy-Efficient Resource Allocation Algorithm Based on Immune Clonal Optimization for Green Cloud Computing". EURASIP Journal on Wireless Communications and Networking, Vol. 64, December 2014.

[15] S. Tayal, "Task Scheduling optimization for the Cloud Computing Systems", International journal of advanced engineering sciences and technologies, Vol No. 5, Issue No. 2, 201, pp. 111-115.

[16] T. Wang, Z. Liu , Y. Chen, Y. Xu, X. Dai, "Load Balancing Task Scheduling based on Genetic Algorithm in Cloud Computing", 12th International Conference on Dependable, Autonomic and Secure Computing, IEEE 2014.

[17] H. Chu, "Service Cost of Resource Scheduling in Cloud Computing based on an Improved Algorithm Combining Support Vector Machine with Genetic Algorithm", International Journal of Grid and Distributed Computing Vol. 9, No. 6 (2016), pp.51-62.

[18] S. Kim, J. Byeon, H. Yu and H. Liu, "Biogeography-Based Optimization for Optimal Job Scheduling in Cloud Computing", Applied Mathematics and Computation, Elsevier, Vol.247, pp. 266-280, 2014.

[19] A.V Lakra, and D. K Yadav, "Multi-Objective Tasks Scheduling Algorithm for Cloud Computing Throughput Optimization", International Conference on Intelligent Computing, Communication & Convergence, 2015.

[20] A. Narwal and S. Dhingra, "Task Scheduling Algorithm Using Multi-Objective Functions for Cloud Computing Environment", International journal of control theory and applications, Vol. 10(14), pp. 227-238, 2017.

[21] N. Bansal, A. Maurya, T. Kumar, et al., "Cost performance of QoS Driven task scheduling in cloud computing". Procedia Computer Science, 57, 2015, pp. 126-130.

[22] M. Abdullahi, M. A. Ngadi, S. M. Abdulhamid, "Symbiotic Organism Search Optimization Based Task Scheduling in Cloud Computing Environment". Future Generation Computer Systems, Vol. 56, March 2016, pp. 640-650.

[23] Y. Sun, J. White, S. Eade, D. C. Schmidt, "ROAR: AQoS-Oriented Modeling Framework for Automated Cloud Resource Allocation and Optimization", Journal of Systems and Software, 2015.

[24] D. G. Feitelson and B. Nitzberg, "Job characteristics of a production parallel scientific workload on the NASA Ames iPSC/860", In: workshop on job scheduling strategies for parallel processing, Springer, 1995, pp. 337–360.

Communication Papers of the Federated Conference on Computer Science and Information Systems pp. 119–122
DOI: 10.15439/2019F297
ISSN 2300-5963 ACSIS, Vol. 20

Building a Clinical Case Reports Database to Facilitate Case-Based Learning and Improve Critical Thinking of Medical Students

Jaroslav Majerník, Peter Jarčuška, Jozef Gonsorčík
Pavol Jozef Šafárik University in Košice, Faculty of Medicine, Trieda SNP 1, Košice, Slovakia
Email: jaroslav.majernik@upjs.sk, peter.jarcuska@upjs.sk, jozef.gonsorcik@upjs.sk

***Abstract*—This paper reports on development of our unified case reports database for medical students. The primary aim was to share professionals' clinical knowledge and skills through one of the most traditional medical experiences sharing form based on modern and anywhere available web services and technologies. Doing so, the clinical case reports that still play an invaluable role in case-based learning and in building of ability of students to think critically, can be used not only to learn and practice theoretical and practical consequences of clinicians' decisions, but also to prepare students for bedside teaching. To develop a generally acceptable system, we had to specify both the structure of education suitable materials and the platform for simple creation and sharing of such electronic outputs. As the result, the database of first clinical case reports was implemented into the information technologies infrastructure at Faculty of medicine in Kosice, Slovakia. It is offered to medical students in the form of web-based portal.**

I. Introduction

CLINICAL experience represents a cornerstone of learning medical and health care practice which allows medical students to understand the high importance of interconnection between theoretical and practical knowledge. Various modern technologies based approaches, including ICT supported problem-based learning, virtual patients etc., were designed and integrated into the curricula of medical education providing institutions. The traditional paper-based patient case reports were often considered a part of an old-fashioned approach. However, they still remained an invaluable source of teaching materials. In particular, this is due to a real clinical background based on true patient stories and integration of professionals' skills and their experiences involved in best available strategies used to solve particular clinical case.

The most of medical education institutions are struggling to ensure a sufficient range of clinical placements for their students. These institutions recognize it is essential and irreplaceable component of medical education. Moreover, the range of clinical learning hours can vary significantly from one country to another [1]. Except of transformation of theoretical knowledge into the practice, the students learn to face stressful situations when they are in clinical learning environment. On the other hand, the learners have to be prepared both the theoretically and pre-clinically before they will be allowed to enter the bedside teaching. Here, the clinical case reports play their education role even though they, as the sources of evidence, are located at the bottom parts of the evidence-based medicine hierarchy [2].

From the educational point of view, the case reports represent wealthy source of practical knowledge and experiences applicable in various medicine oriented pedagogical processes. Basically, the clinical case reports are presented in the form of clinical observations documentation that describe details of common and rare cases; characteristics of known and unknown diseases; novelty and/or new ideas in medicine; variations in diseases and their combinations; positive and negative effects of interventions; side effects of drugs usage as well as the ways professionals use to solve individual patient health related problems, respecting their best recent knowledge.

Because of the absence of general reporting guidelines designed for case reports, the clinicians and teachers report their clinical cases to the students in various heterogeneous forms using either electronic or paper-based approach. On the other hand, there are efforts of several medical associations to develop guidelines for clinical case reports in specific clinical domains. A positive example of such developments in this area is the initiative of international group of experts that developed Case Reports guidelines (CARE guidelines) [3, 4]. Aiming on reduction of bias, increasing of transparency and providing of first outputs of clinically applicable methods, these guidelines are useful mostly for publication of high quality clinical cases in scientific journals. However, many of the medical journals already stopped to publish clinical case reports as for the low citation index and thus negative effect on journal's impact factor [5]. In addition, the contemporary studies pointed out

This work was not supported by any organization

©2019, PTI

the case reports based on guidelines, whether modified or derived from CARE guidelines, have their usefulness and they are not rarely cited. Therefore, the well-written clinical case reports are again popularly published in newly established medical journals or in special supplements, volumes or websites of many prestigious publishers. From the educational point of view focused on higher education levels, the reporting of modern clinical cases tends to take a form of narrative stories that reveal gradually a solution of clinical problem [6, 7].

CARE guidelines have the potential to be a framework of clinical case reports for education purpose, but one has to be aware of broad variety of both general as well as specific objectives given in individual clinical disciplines that cannot result in a single case report framework applicable to all clinical cases and domains. As the evidence, various protocols and architectures of clinical case reports were published across different clinical disciplines [8, 9] or in biomedical applications [10, 11, 12]. This suggest the idea that the case reports significantly support clinical reasoning while forming an interface between medical knowledge and medical practice [13].

The inconsistency in case reports based education materials offered to medical students at our faculty forced us to initiate activities leading to the development of clinical case reports database and to the motivation of our academic clinicians to participate on it and to utilize it in education process. In contrast to the above-mentioned publishing phenomena of clinical case reports, and because of the educational background of our work, we faced an opposite role. Our main task was to design not discipline specific, but rather generalized case report framework to suit as many disciplines as possible while keeping authors a reasonable space for integration of their specifics that may vary from case to case. Also, this interdisciplinary acceptable framework should be offered to the teachers together with online available tools used to design and share individual elements of their clinical cases to their learners.

II. Materials and Methods

Before the system of clinical case reports was developed and integrated into the teaching process, we had to solve both the methodological as well as the technological problems. The methodological problems were aimed to find a uniform framework of case reports that will fit to the most of the clinical disciplines taught at the faculty. To solve the technologies related problems we searched for the best way, which our teachers will use to create, to share and to maintain the database of clinical case reports and that will be available to learners anytime and anywhere. Also, this repository had to be developed in the way, which will serve medical students as additional source of clinically relevant study material and which should help them in better preparation for their later real clinical praxis.

Individual forms of case reports presented to our students by the teachers at individual clinical departments were analysed in order to identify the best consensus across various clinical disciplines. Not surprisingly, the clinical teachers used a broad scale of options, starting by reading the patient history, examination notes, laboratory test results etc. in paper based health records, through PowerPoint like presentations and ending with the study of the records and laboratory test results stored in clinicians' information systems. However, no one offered comprehensive and structured report prepared in compliance with CARE guidelines or similar ones. In addition, no of the case reports was available outside of the university network even in secured platforms. Personal discussions followed afterwards and the committee consisting of vice-deans and guarantors of particular clinical study fields reviewed and annotated importance and usefulness of individual sections in these forms of case reports. The conclusion was to use a generalized structure of education suitable case reports as it was presented in [14]. The proposed generalised structure of clinical case reports consists of following elements:

1) Title - topic of interest related to the patient, disease, symptoms, interventions etc.;
2) Author(s) details - author(s) name(s) and affiliation to department(s) of medical faculty;
3) Medical discipline(s) - assigned medical discipline(s) related to the problem to be solved;
4) Annotation image - image associated to the problem of presented case report;
5) Annotation - a brief introduction to the problem and/or short summary of the case report background;
6) Patient's history (anamnesis) - medical, social and family history, abuses, current health problems and symptoms etc.;
7) Laboratory tests results - clinical findings, results of relevant physical examinations, special laboratory tests etc.;
8) Imaging methods - imaging diagnostic methods where applicable, images from modalities used, description of findings and obtained results;
9) Diagnosis - main diagnosis specified according to ICD-10 and resulting from clinical findings and tests, used also to search for similar case reports published at faculty's clinical case reports database/portal;
10) Differential diagnosis - diagnostic reasoning with all considered diagnoses in patient assessment process;
11) Therapy - therapeutic interventions performed, important dates, treatment administration, prognostic details etc.;
12) Discussion and comments - results of the case, assessment of the results achieved, recent patient's status, strengths of the case, risks and limitations if any, recommendations of relevant literature;
13) Keywords - key elements of the case, used also to search for similar case reports published at faculty's clinical case reports database/portal and
14) Courses - a list of associated courses taught at the faculty in medical disciplines, used also to search for another

education materials and articles related to the course(s) published at faculty's clinical case reports database/portal.

Bearing in mind all the recent possibilities of modern technologies and the fact that our teachers presented clinical cases to our students in various heterogeneous forms, including paper-based and oral presentations during face-to-face teaching activities, we decided to focus their attention and pedagogical activities in this area to the one unified and online available place. Thanks to the MEFANET network and its ICT activities that already brought several interdisciplinary useful educational tools [15, 16], we found the solution of this problem in tools used within this educational network. In addition, we wanted to maintain the nature of traditional case reports and not to transform them into the simulation or standalone learning management systems as it was done for example in [17].

As the technological solution, we decided to use our local instance of MEFANET's portal platform to host the repository of our clinical case reports developed for various clinical disciplines. This decision resulted from great potential of the platform to manage multimedia content as well as to provide different user roles and access to the published materials. The advantages for teachers include the possibility to use generalized structure of case reports, minimal requirements on their technical skills and many others. On the other hand, the students can find everything in one system and they can study individual clinical cases, together with other types of study materials, wherever and whenever they need. Furthermore, using this way and depending on authors decision, the individual case reports can be exported to the MEFANET's Central Gate and thus shared to the learners studying at all medical faculties in Czech Republic and Slovakia.

III. Results

To ensure as extensive impact of our work as possible, to start our project efficiently, to find the best consensus for all disciplines, to address the widest community of clinical teachers and finally to reach continuously growing number of clinical case reports, the management of the faculty organized a meeting with academic clinicians and teachers in September 2017. More than 60 clinicians took part in this meeting and they were informed about ideas and aims of that educational activity. Case Report framework was presented to the participants and consequently it was revised together with opened discussion to prove it is acceptable for the academics teaching medical students of our faculty. This also led to elimination of potential confusions and incompleteness of some parts or descriptions. Then, the participants were informed about the Portal of multimedia support in the education of clinical and health care disciplines at Faculty of Medicine at Pavol Jozef Safarik University in Kosice as well as about the ways used to publish clinical case reports on this Portal. The 14 above-mentioned sections of case report structure related to the portal presentation were approved.

The core idea of our initiative was accepted quite positively and most of the participants agreed to participate on first round of case reports development that was opened from 1st October 2017 to 31st December 2017. The first case reports were send to the redaction within few days after the meeting took place and once the post-meeting discussions were finished. As we expected, the majority of the case reports was prepared in internal medicine, surgery and dental medicine too. The academic clinicians generated several tenths of cases within this three-month period. 63 of them were finished, rearranged to fit the portal's framework and layout and shared to the students.

The list of annotations shared at the web-based portal is shown in Figure 1. The Figure 2 shows an example of open clinical case report with access to the multimedia content.

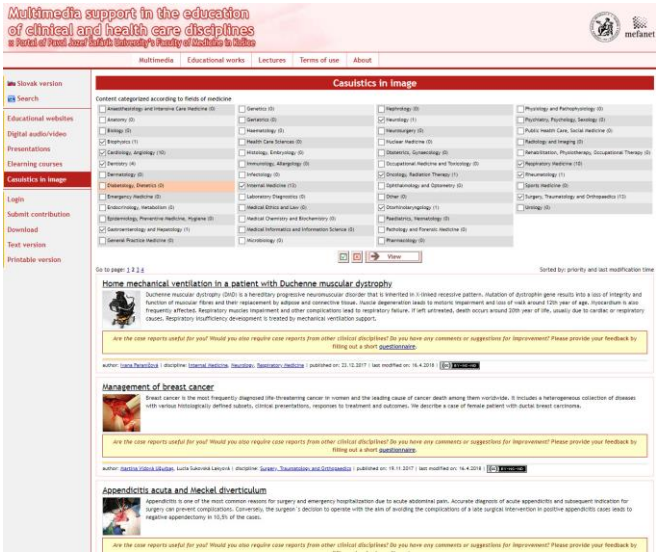


Fig. 1 The structure of annotations shared as clinical case reports database at the Portal of multimedia support in the education of clinical and health care disciplines at Faculty of Medicine at Pavol Jozef Safarik University in Kosice.

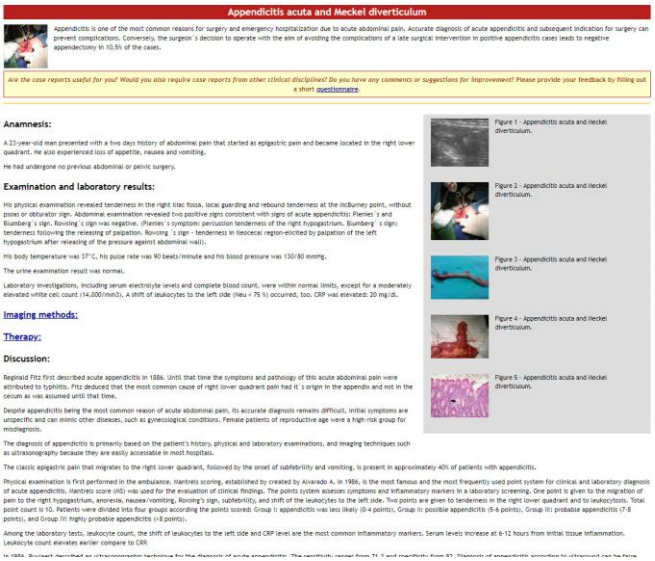


Fig. 2 An example of clinical case report published at faculty's portal.

Then the activity to create clinical case reports database at our faculty continued and now the authors create their new and improve their existing cease reports depending on the most recent knowledge and students' feedback too.

There were totally 76 case reports finished and published at the Portal until now. All the case reports are published according to the portal's rules. To ensure the highest quality, guarantors of particular clinical disciplines review the content of all case reports. Then, each reviewed case report is labelled by the review logo to allow students recognizing reviewed and unreviewed clinical case reports. The reviewed case reports are automatically send to the Central Gate of MEFANET's portals, thus the students of all medical faculties in Czech republic and Slovakia can find them in one place and use them in their studies no matter which faculty they are from.

Another added value of our work, as we suppose, is that we expect increased motivation to create clinical case reports for education also in academic clinicians from another medical faculties involved in MEFANET. The motivation can be increased not only because of our pilot round of development, but maybe also because of the first positive students' feedbacks we noticed after publishing our pilot case reports.

IV. Conclusion

Even if the scientific impact of clinical reports is considered as very low, we suggest the critical thinking of medical students can be improved thanks to the proper combination of high-quality clinical case reports, other traditional and electronic education methods used at the faculty and clinical bedside teaching. Except of others, the benefits for medical students included access to the valuable clinical experiences.

Medical education involves various teaching methods and strategies showing students clinical stories and examples based on real patients' health related problems. Even if the evidence-based medicine is considered to be one of the most relevant methods revealing background of systematic research and scientific clinical outputs, individual case reports have still a great potential to increase students' medical knowledge. Finally, we have to conclude the case reports offer students very challenging opportunities to evolve their clinical skills, but also to foster scientific and critical thinking.

Acknowledgment

Results presented in this work were obtained with the support of the national agency's grant KEGA 011UPJS-4/2019 "Increasing of competences and critical thinking level in students of medical study programs using simulation tools of Problem-Based Learning and Evidence-Based Medicine".

References

[1] Liu M., Chan Y.M., Tee S., Gu K., Luo Z.M., Wong T.K.S., "An iterative approach to enhance the clinical learning experience in Macao nursing education", International Journal of Nursing Sciences, 6, 2019, pp. 216-220, http://dx.doi.org/10.1016/j.ijnss.2019.01.005.

[2] van Haselen R.A., "Towards improving the reporting quality of clinical case reports in complementary medicine: Assessing and illustrating the need for guideline development", Complementary Therapies in Medicine, 23, 2015, pp. 141-148, http://dx.doi.org/10.1016/j.ctim.2015.01.009.

[3] CARE guidelines. [On-line] Available at WWW: <http://www.care-statement.org>

[4] Gagnier J.J., Kienle G., Altman D.G., Moher D., Sox H. Riley D., "The CARE Guidelines: Consensus-based Clinical Case Reporting Guideline Development", Global advances in health and medicine. 2(5), 2013, pp. 38-43.

[5] Firat A.C., Araz C., Kayhan Z.. "Case reports: Should we do away with them?", Journal of Clinical Anesthesia. 37, 2017, pp. 74-76, http://dx.doi.org/10.1016/j.jclinane.2016.10.006.

[6] Hurwitz B., "Narrative constructs in modern clinical case reporting", Studies in History and Philosophy of Science, 62, 2017, pp. 65-73, http://dx.doi.org/10.1016/j.shpsa.2017.03.004.

[7] Agha R.A., Fowler A.J., Saetta A., Barai I., Rajmohan S., Orgill D.P. "The SCARE Statement: Consensus-based surgical case report guidelines", International Journal of Surgery, 34, 2016, pp. 180-186, http://dx.doi.org/10.1016/j.ijsu.2016.08.014.

[8] Agha R.A., Farwana R., Borrelli M.R., Tickunas T., Kusu-Orkar T., Millip M.C., Thavayogan R., Garner J., Orgill D.P., "Impact of the SCARE guideline on the reporting of surgical case reports: A before and after study", International Journal of Surgery, 45, 2017, pp. 144-148, http://dx.doi.org/10.1016/j.ijsu.2017.07.099.

[9] van Haselen R.A., "Homeopathic clinical case reports: Development of a supplement (HOM-CASE) to the CARE clinical case reporting guideline", Complementary Therapies in Medicine, 25, 2016, pp. 78-85, http://dx.doi.org/10.1016/j.ctim.2015.12.019.

[10] Rajťúková V., Poláček I., Tóth T., Živčák J., Ižaríková G., Kovačevič M., Somoš A. Hudák R., "The manufacturing precision of dental crowns by two different methods is comparable", Lekar a Technika, 46(4), 2016, pp. 102-106.

[11] Kubicek J., Rehacek T., Penhaker M., Bryjova I., "Software simulation of CT reconstructions and artifacts", Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST, Volume 165, 2016, Pages 428-437, http://dx.doi.org/10.1007/978-3-319-29236-6_41.

[12] Riley D.S. et al., "CARE guidelines for case reports: explanation and elaboration document", Journal of Clinical Epidemiology,89, 2017, pp. 218-235, http://dx.doi.org/10.1016/j.jclinepi.2017.04.026.

[13] Linsena A., Elshout G., Pols D., Zwaan L., Mamede S., "Education in Clinical Reasoning: An Experimental Study on Strategies to Foster Novice Medical Students' Engagement in Learning Activities", Health Professions Education, 4, 2018, pp. 86–96, http://dx.doi.org/10.1016/j.hpe.2017.03.003.

[14] Majernik J., Gonsorcik J. Repository of Clinical Case Reports. Mefanet 2017, Brno, 2017; 26-27.

[15] Schwarz D., Štourač P., Komenda M., Harazim H., Kosinová M., Gregor J., Hůlek R., Smékalová O., Křikava I., Štoudek R., Dušek L., "Interactive algorithms for teaching and learning acute medicine in the Network of Medical Faculties MEFANET", Journal of Medical Internet Research, 15 (7), 2013, pp. 135, http://dx.doi.org/10.2196/jmir. 2590.

[16] Komenda M. et al., "Curriculum Mapping with Academic Analytics in Medical and Healthcare Education", PLoS ONE, 10 (12), 2015, e0143748, http://dx.doi.org/10.1371/journal.pone.0143748.

[17] Ali M. et al., "iCBLS: An interactive case-based learning system for medical education. International Journal of Medical Informatics", 109, 2018, pp. 55-69, http://dx.doi.org/10.1016/j.ijmedinf.2017.11.004.

25th Conference on Knowledge Acquisition and Management

KNOWLEDGE 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

Communication Papers of the Federated Conference on Computer Science and Information Systems pp. 125–129
DOI: 10.15439/2019F196
ISSN 2300-5963 ACSIS, Vol. 20

Innovative Modular Approach based on Vehicle Routing Problem and Ant Colony Optimization for Order Splitting in Real Warehouses

Emir Žunić[1,3], Sead Delalić[2,1], Zlatan Tucaković[3,1], Kerim Hodžić[3,1], Admir Beširević[2,1]
[1]Info Studio d.o.o., Sarajevo, Bosnia and Herzegovina
[2]Faculty of Science at University of Sarajevo, Bosnia and Herzegovina
[3]Faculty of Electrical Engineering at University of Sarajevo, Bosnia and Herzegovina
Emails: emir.zunic@infostudio.ba, {delalic.sead, besirevic.admir}@pmf.unsa.ba, {zlatan.tucakovic, kerim.hodzic}@etf.unsa.ba

***Abstract*—A crucial part to any warehouse workflow is the process of order picking. Orders can significantly vary in the number of items, mass, volume and the total path needed to collect all the items. Some orders can be picked by just one worker, while others are required to be split up and shrunk down, so that they can be assigned to multiple workers. This paper describes the complete process of optimal order splitting. The process consists of evaluating if a given order requires to be split, determining the number of orders it needs to be split into, assigning items for every worker and optimizing the order picking routes. The complete order splitting process can be used both with and without the logistic data (mass and volume), but having logistic data improves the accuracy. Final step of the algorithm is reduction to Vehicle Routing Problem where the total number of vehicles is known beforehand. The process described in this paper is implemented in some of the largest warehouses in Bosnia and Herzegovina.**

I. Introduction

THE workflow of large distribution companies consists of acquiring items from the manufacturer and storing them in large warehouses. When orders are made, items are delivered and distributed to customers. Distribution companies can deliver thousands of orders daily to various types of customers: smaller markets, shopping malls, restaurants etc. Each order has to be collected from the warehouse.

The process of collecting orders is usually called the order picking and it takes a considerable part of the warehouse workflow. The order picking is the most demanding process and it takes approximately 55% of the warehouse operating cost [1]. It is usually done by warehouse workers who are walking between the racks and collecting items in a trolley or a cart. In some cases items are picked up with a forklift. The picking route is determined before the process of collecting items begins. Each process is controlled by the Warehouse Management System (WMS) [2].

The process of finding optimal route is equivalent to the standard Traveling Salesman Problem (TSP) with additional constraints such as picking heavy objects first and placing them at the bottom cart or placing items last because they have special temperature requirements. Many heuristic algorithms are given to solve this problem, so the process of order picking can be optimized.

Orders can contain between one and more than a thousand of distinct items and in that case, the picking process can be a challenging task. The optimal route can have the length of several kilometers and the mass and the volume of picked items can exceed the capacity one worker can handle. In that case, the worker must pause the picking process, go to the unloading location and unload items, prior to resuming the order picking process. The problem with this solution are common cases in which the unloading location is far from the current picking location, so the route length is significantly increased. However, the order splitting is imposed as an intuitive and natural solution. The process of order splitting can improve order picking process in general, but the order splitting has to be done in a way that the operating costs are reduced. This paper presents the innovative approach for order splitting implemented in real warehouses.

The main goal is to split the order to a number of smaller and balanced orders. This can also reduce a psychological burden on the worker, and increase the possibility of establishing a better balance between tasks among the workers. The orders can be balanced by the total route length, mass, volume or other parameters. The mass, volume or other logistic data are not always available or accurate. The process of order splitting needs to be adjusted in a way that works well in both cases.

The process of order splitting consists of several steps. The first step is to determine the number of smaller orders which the main will be divided into. Some orders do not need to be divided, because they do not require the transportation over a large distance and do not contain volume and mass that overload one worker. The second step is the process of dividing items to smaller orders and the process of route optimization for each order.

The algorithm described in this paper is implemented in some of the largest warehouses in Bosnia and Herzegovina. These warehouses deliver in real-time, so the orders are picked as soon as possible. In warehouses where the 48-hours delivery

©2019, PTI

system is used, orders are known at the beginning of the day. This means that everything can be planned before the first order is collected. In these cases, order batching can improve the picking process. The order batching is the process of merging smaller orders into a larger one, so the worker can simultaneously pick all of them. The batching is not possible when the orders are not known before, which is the case in real-time delivery systems. The algorithm presented in this paper can be of great importance for advanced order splitting in warehouses with real-time delivery system and can be used in warehouses with non-standard layouts.

In the first section, the short introduction and motivation for this paper is given. The literature review is given in the second part. In the third section, the complete process of order splitting is described. Results are given in the fourth section, as well as the comparison of the proposed algorithm and other popular approaches. Last section gives the short summary and the future work ideas.

II. Related Work

The algorithm implemented in the paper is an upgrade of the smart Warehouse Management System described in paper [3]. Order splitting requires distance calculation between any two positions in a warehouse. System uses distance calculation inside a warehouse using a dynamical algorithm for traditional warehouse layouts implemented in paper [4], as well as for nontraditional layouts, described in paper [5]. Order picking route can be additionally reduced if items are placed according to their ordering amount and frequency. More frequent items are at better locations, i.e. closer to the exits. This is implemented as algorithm and described in [6]. Orders are picked faster if workers have enough items in warehouse picking zone and they do not have to use forklifts. Algorithm that predicts item sales based on sales history is described in [7].

The algorithm proposed in this paper uses similarity of order splitting and Vehicle Routing Problem (VRP), while criteria is that routes should have a balanced load. Task of finding the optimal solution for VRP is extremely hard, but sub-optimal solutions can be found using some heuristic or metaheuristic algorithms. Variant of VRP with balanced load routes is presented in [8]. Tabu search is one heuristic method that gave some of the best solutions in this field. In [9] author described approximate methods based on descent, hybrid simulated annealing/tabu search with various search methods. Algorithm computation time is optimised with specific data tables organisation. Another tabu search based on heuristic is TABUROUTE and it is described in paper [10]. It is based on a row of adjacent solutions acquired by generalized insertion procedure which is replacing one vertex from route to route and recalculating route lengths. Customer window constraints represent period of time in which customer is available to receive transported goods. Those constraints are common in real life situations and they can make VRP problem solving more complicated. They are usually referred as VRPTW (Vehicle Routing Problem with Time Windows) in literature. Mathematical model and and optimization-based algorithms using route length minimization as criteria is described in paper [11]. Operational research methods can also help solving this type of problems. In paper [12] optimization algorithm based on linear and dynamic programming with feasible column generation is proposed. Linear programming solution was used as lower bound for branch-and-bound algorithm used for solving integer set partitioning formulation. This concept delivered optimal results for a hundred customers VRP problems. Algorithm and concept described in this paper uses Ant Colony Optimization for solving VRP. Ant Colony Optimization is introduced and described in paper [13]. That algorithm gives very good sub-optimal solution, within 1% of known optimal solutions, and uses multiple ant colonies. Another approach for order splitting based on Poisson processes is given in [14], assuming that arrival process is distributed by Poisson distribution.

III. Case Study

In this section, the complete algorithm for order splitting is described. The main goal of the algorithm is to optimally split the given order to smaller ones.

The proposed algorithm consists of three main steps:

- Collecting data and distances calculating
- Determining the number of smaller orders
- Order splitting

A. Collecting data and distance calculating

The process of collecting data begins with determining locations in warehouse from which the items will be collected. In the concept of smart warehouse the items with the earliest expiration date are picked, so the locations are easy to determine. As mentioned before, many warehouses do not have complete logistic data. Often, the mass or the volume of the items are missing. In addition, in many practical cases, the data are inaccurate (inconsistent unit of measure or typographical error). Therefore, the proposed algorithm for order splitting must work smoothly in situations with inaccessible logistic data. The following data are observed: item location, number of ordered units, mass and volume (if known). Before the next step of algorithm, individual item records are aggregated for each location. The items will be referred as locations in further discussion. The distance between each two locations are calculated. Papers [4] and [5] propose the algorithm for distance calculation in warehouses of various layouts. The mentioned algorithm is used in this implementation.

B. Determining the number of smaller orders

The process of calculating number of smaller orders (k) consists of checking the size of the order and using known data for the split. As size metric total number of different items, the total number of items, the total mass (m_{tot}) or the total volume of items (V_{tot}) can be used, respectively. Known data include: total distance of the optimal order picking route (d_{tot}), maximal mass (m_{max}) and volume (V_{max}) one worker can take, total number of available workers, maximal distance

one worker should travel (d_{max}), etc. Total order picking route distance d_{tot} is a solution of TSP, and it represents the distance one worker must travel to complete whole order. This distance does not need to be optimal, since for the sake of the algorithm only good sub-optimal solution is needed. Other data can vary from one warehouse to another, and they can be acquired through conversation with warehouse personnel or from database. Finally, the number of smaller orders is calculated as (1).

$$k = \left\lceil \max \left\{ c_d \cdot \frac{d_{tot}}{d_{max}}, c_m \cdot \frac{m_{tot}}{m_{max}}, c_V \cdot \frac{V_{tot}}{V_{max}} \right\} \right\rceil \quad (1)$$

Parameters c_d, c_m and c_V are correction coefficients, and have non-negative values. For example, if $c_m = 0$, then masses are not important in calculation of k. This is useful if masses of items are not available. It might be good to set coefficient at 1.1 to 1.3, to gain some flexibility in route calculations.

C. Order splitting

When the optimal number of smaller order is decided, it is necessary to solve two problems: distribution of items in smaller orders and route optimization for each of them.

Mathematically, the problem can be described as follows. Let n be the number of different items locations denoted by $a_1, a_2, \ldots, a_n$. As stated, the calculated number of smaller orders is denoted by k. The distance between locations a_i and a_j is given by $d_{i,j}$, for all $i, j = \overline{1,n}$. Aggregated mass and volume for items on location a_i are denoted by a_i^m and a_i^V, for all $i = \overline{1,n}$.

The algorithm distributes the items in k by pairs disjoint arrays $(a_{i,1}, a_{i,2}, \ldots, a_{i,l_i})$, where l_i denotes the number of locations in i-th route, $i = \overline{1,k}$. Item $a_{i,j+1}$ is collected just after the item $a_{i,j}$, for all $i = \overline{1,k}$ and $j = \overline{1, l_i - 1}$. The starting and ending locations are same for each smaller order because workers have to take the order at the starting location and merge the orders at the end.

Before the process of optimization begins the objective function has to be determined. In practice, the goal is to optimize routes in such a way that the workload of each route is balanced, but at the same time the overall length is as low as possible. A large number of orders can not be optimized in a way that the mass, volume and route length are simultaneously balanced. Therefore, the proposed objective function that is going to be minimized is stated as:

$$D + \max_{i=\overline{1,k}} (c_m \cdot m_i + c_V \cdot V_i + c_d \cdot d_i), \quad (2)$$

where:

- D is the total length of all routes,
- m_i is the total mass for the route i,
- V_i is the total volume for the route i,
- d_i is the total length of route i.

Constants c_m, c_V and c_d are set depending on the needs of the warehouse and the desired criteria. When logistic data are not available, parameter c_m or c_V should be set to 0.

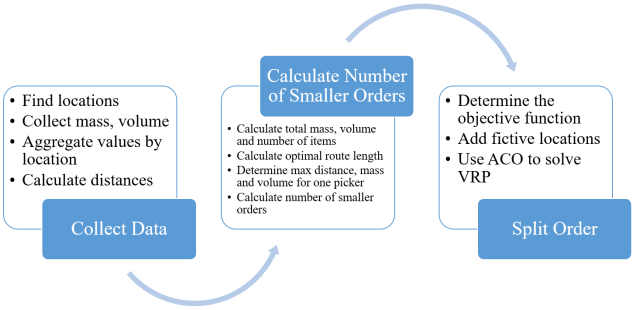


Figure 1. Proposed algorithm

The complete process of items distribution and route optimization is equivalent to the process of solving VRP. The equivalence of customers locations and items locations are established, as is the number of routes with the number of vehicles available. The only difference is the objective function where the goal is to find the balanced routes.

In this paper, the VRP is solved by solving the TSP. The VRP is transformed to TSP by adding starting and ending locations for each route as customers to TSP. Each distance between two added locations is set to $+\infty$, so we can guarantee that there will be no optimal route where the added customers are adjacent. The TSP is solved using the Ant Colony Optimization (ACO) [15]. The complete process is shown in Figure 1.

IV. Results

This section provides a comparison between the proposed algorithm and two approaches that are commonly used in practice for large orders.

1) One worker: In many warehouses, the orders are not split despite their size. In case of splitting, the total distance can not be shorter than in the case when one worker is picking the complete order. Meanwhile, orders can have large mass and volume that are greater than the worker capacity. Workers are forced to pause the picking process and unload the picked items on the final location before continuing with the order picking process. In those cases, the traveled distance can be shorter in orders that are split. The main advantage of this approach is the fact that it is not necessary to hire new workers to merge split orders. On the other hand, in many warehouses, some workers are responsible for checking the collected orders and the merging process can be addressed to them.

2) Divide by sectors: In practice, the common way of solving the order splitting problem is to divide the warehouse into sectors. The sectors are statically determined, usually by articles type or brand. Then, each sector has a worker in charge and the route is optimized only for that sector. This approach has several drawbacks. The main problem with this approach are orders where majority of locations and items are in one sector. The employee who collects the order in mentioned sector can be overburdened while other employees are spared and their routes are not profitable as they collect

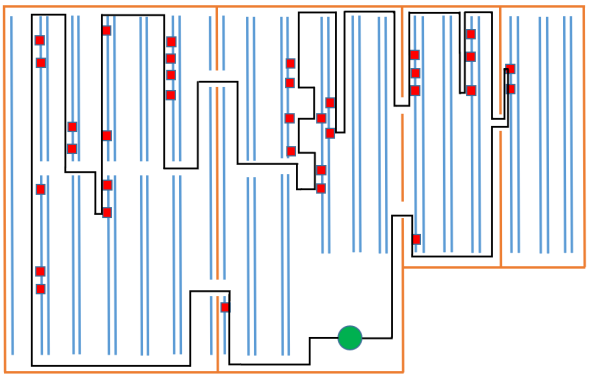


Figure 2. The optimal route for one worker and locations of items

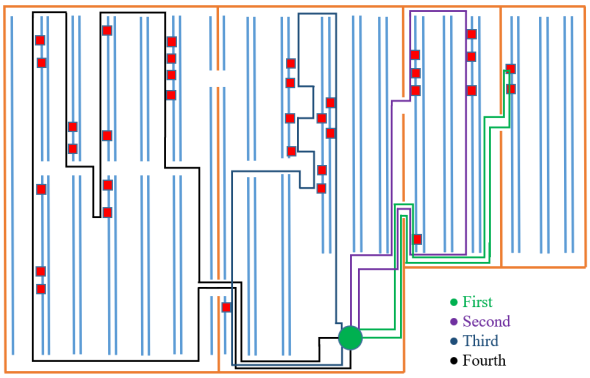


Figure 4. Order divided by sectors

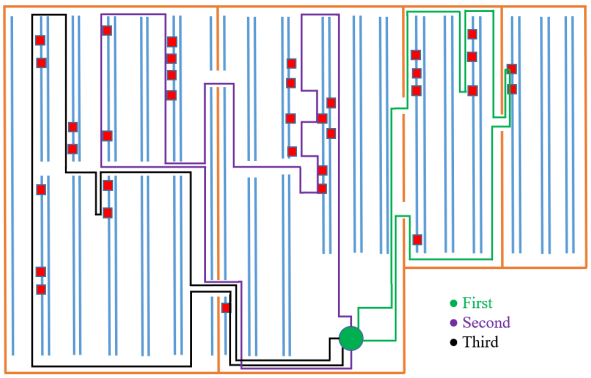


Figure 3. Order with overloads

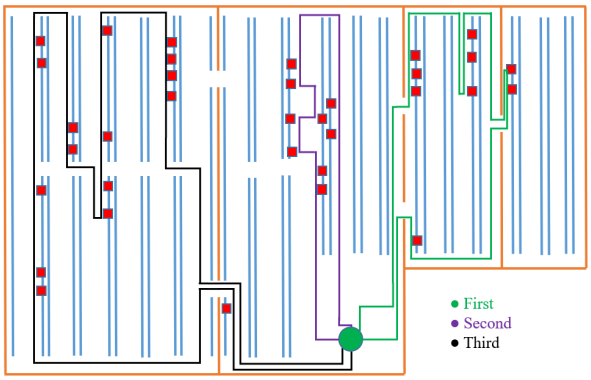


Figure 5. Order given by the proposed algorithm

a small number of items. This approach of order splitting presents the entire warehouse as several smaller warehouses. Workers are better acquainted with their environment and product positions, which is one of the biggest advantages. The sectors are most often created according to the type of goods such as food, animal food, cosmetics, frozen goods etc.

Recognized advantages of the proposed algorithm will be shown on the example of one real-world order. All 34 locations and complete order picking route are shown in Figure 2.

The optimal route length for one worker to complete the order is $d_{tot} = 1130$ meters. It is important to state that the route was longer than the predetermined optimal route length per worker d_{max}. It is also observed that the worker reached maximum capacity during the process twice, paused picking and went to unloading location to unload. Due to this action, the route length is increased by 390 meters. Taking this in account, the route with three smaller routes is shown in Figure 3. All constants were determined by the warehouse manager.

When the observed order was split on routes based on the sectors as shown in Figure 4, routes were not balanced. Each room of the presented warehouse was one sector. The total length was 1880 meters. The longest route was 2.75 times longer than the shortest one and the mass was 450% larger.

In the Figure 5 the order splitting by proposed algorithm is given. The total route length was 1430 meters. The longest route was 2.5 times longer than the shortest one, but the mass was 14% larger. As we can see, the proposed routes are well balanced and the algorithm can solve the drawbacks of earlier approaches. The problems with pausing the order picking process and large distances walked by one worker are solved, as is the problem of unbalanced orders in sectors. At the same time, the total length is not significantly increased.

In addition to the order from the example, the algorithm was tested in one of the largest warehouses in Bosnia and Herzegovina on 1,000 large orders (twenty or more picking locations where one location can contain many items). When warehouse started operating, the orders were not split. Later, the algorithm for splitting to sectors and the proposed algorithm for order splitting were implemented. The same orders were tested using all three mentioned approaches. All 1,000 orders are randomly selected larger orders where logistic data are fully available for easier comparison.

The average number of locations is 34.37 and the average mass is 236.51 kilograms. After solving the TSP and calculating the route length for the TSP, 729 orders exceeded the predetermined limit for route length for one worker. At the same time, 545 orders exceeded mass and volume limit. Detailed results are shown in Table I.

The main observed advantage of this algorithm is the dynamical sector creation and simultaneous reduction worker

Table I
COMPARISON OF ALGORITHMS FOR 1,000 REAL-WORLD ORDERS IN VARIOUS WAREHOUSE LAYOUTS.

Average	**No split**	**No split (logistic)**	**Sector**	**Algorithm**	**Algorithm (logistic)**
Number of smaller routes	1	2.92	4	2.57	2.83
Total distance (m)	1240	1780	2118	1330	1592
Longest route distance (m)	1240	1212	1190	540	582
Maximum route mass (kg)	236.51	126.39	168.22	125	118.23
Maximum route volume (m^3)	1.72	1.12	1.32	1.12	1.22
Number of times route length exceeded optimal length d_{max}	728	542	280	0	0
Walking time to complete order (parallel picking) (min)	13.52	21.79	12.98	5.89	6.32

load when necessary. This combines the good features of the practice approaches, but at the same time solves their biggest drawbacks. Described algorithm can significantly increase the delivery time if smaller orders are picked at the same time. As stated in [6], most of the orders consist out of 10 or less different items. This means that good positioning of those items can transform the warehouse into smaller one for many orders. In the combination with the proposed algorithm, the maximum distance constant will be rarely exceeded.

The proposed algorithm can be used in a combination with an algorithm for dividing into sectors, where for each smaller order the main sector is determined. The main sector can be determined as the sector where most locations from that order are placed or as the sector where the largest part of the picking route is located. In this way, a primary sector can be assigned to each worker. Then, every order is assigned to a worker with the same primary sector. Each worker can pick some items from locations outside the sector, but the main route will be in the primary sector. This way, the route length is not increasing, but each worker is picking orders in one part of the warehouse, which is emphasized as an advantage of dividing orders by sectors.

V. CONCLUSION AND FUTURE WORK

In this paper, the motive for order splitting is described, and the complete process of optimal order splitting is presented. The first step was evaluation whether the given order should be split or not, the second was calculating the optimal number of smaller orders it needs to be split into and the final step was assigning items to those smaller orders and optimizing the route for each of the orders. This is implemented as part of the smart WMS in some of the largest warehouses in Bosnia and Herzegovina. The results are satisfying, as well as feedback from warehouse management and workers.

Future research will focus on improving the algorithm with different optimization method. Alongside that, future work will be focused on upgrading mentioned Warehouse Management System with additional features. First of them will be order batching - order picking of many smaller orders or parts of larger orders during one visit, which could be useful for warehouses with 48 hours delivery time, since it cannot be used in real-time.

REFERENCES

[1] D. M. H. Chiang, C. P. Lin and M. C. Chen, "The adaptive approach for storage assignment by mining data of warehouse management system for distribution centres," *Enterprise Information Systems 5.2*, 2011. https://doi.org/10.1080/17517575.2010.537784

[2] R. de Koster, T. Le-Duc, and K. J. Roodbergen, "Design and control of warehouse order picking: A literature review," *Eur. J. Oper. Res.*, 2007. https://doi.org/10.1016/j.ejor.2006.07.009

[3] E. Zunic, S. Delalic, K. Hodzic, A. Besirevic, and H. Hindija, "Smart Warehouse Management System Concept with Implementation," in *2018 14th Symposium on Neural Networks and Applications, NEUREL 2018*, 2018. https://doi.org/10.1109/NEUREL.2018.8587004

[4] E. Zunic, A. Besirevic, R. Skrobo, H. Hasic, K. Hodzic, and A. Djedovic, "Design of optimization system for warehouse order picking in real environment," in *ICAT 2017 - 26th International Conference on Information, Communication and Automation Technologies, Proceedings*, 2017. https://doi.org/10.1109/ICAT.2017.8171630

[5] E. Zunic, A. Besirevic, S. Delalic, K. Hodzic, and H. Hasic, "A generic approach for order picking optimization process in different warehouse layouts," in *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2018 - Proceedings*, 2018. https://doi.org/10.23919/MIPRO.2018.8400183

[6] E. Zunic, H. Hasic, K. Hodzic, S. Delalic, and A. Besirevic, "Predictive analysis based approach for optimal warehouse product positioning," in *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2018 - Proceedings*, 2018. https://doi.org/10.23919/MIPRO.2018.8400174

[7] E. Zunic, K. Hodzic, H. Hasic, R. Skrobo, A. Besirevic, and D. Donko, "Application of advanced analysis and predictive algorithm for warehouse picking zone capacity and content prediction," in *ICAT 2017 - 26th International Conference on Information, Communication and Automation Technologies, Proceedings*, 2017. https://doi.org/10.1109/ICAT.2017.8171629

[8] M. N. Kritikos and G. Ioannou, "The balanced cargo vehicle routing problem with time windows," *Int. J. Prod. Econ.*, 2010. https://doi.org/10.1016/j.ijpe.2009.07.006

[9] I. H. Osman, "Metastrategy simulated annealing and tabu search algorithms for the vehicle routing problem," *Annals of operations research 41.4*, 1993. https://doi.org/10.1007/BF02023004

[10] M. Gendreau, A. Hertz, and G. Laporte, "A tabu search heuristic for the vehicle routing problem." *Management science 40.10*, 1994. https://doi.org/10.1287/mnsc.40.10.1276

[11] S. Acharya, "Vehicle Routing and Scheduling Problems with time window constraints-Optimization Based Models," *Int Jr. of Mathematical Sciences Applications 3.1*, 2013.

[12] M. Desrochers, J. Desrosiers and M. Solomon, "A new optimization algorithm for the vehicle routing problem with time windows," *Operations research 40.2*, 1992. https://doi.org/10.1287/opre.40.2.342

[13] J. E. Bell and P. R. McMullen, "Ant colony optimization techniques for the vehicle routing problem," *Adv. Eng. Informatics*, 2004. https://doi.org/10.1016/j.aei.2004.07.001

[14] M. Yu, "Enhancing Warehouse Performance by Efficient Order Picking," 2008. https://doi.org/10.1287/opre.37.3.404

[15] M. Dorigo and L. M. Gambardella, "Ant colonies for the travelling salesman problem," *BioSystems*, 1997. https://doi.org/10.1016/S0303-2647(97)01708-5

Joint 39th IEEE Software Engineering Workshop and 6th International Workshop on Cyber-Physical Systems

THE 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 25th 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 31st 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 Ameur, Yamine,** IRIT/INPT-ENSEEIHT, 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 Dennmark, 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

Communication Papers of the Federated Conference on Computer Science and Information Systems pp. 133–141
DOI: 10.15439/2019F253
ISSN 2300-5963 ACSIS, Vol. 20

A roadmap for a comparison framework for an adaptable software process improvement framework in small software companies

Micheal Tuape
University of Botswana
Department of Computer Science
Gaborone Botswana
tuapem@ub.ac.bw

Yirsaw Ayalew
University of Botswana
Department of Computer Science
Gaborone Botswana
ayalew@mopipi.ub.bw

Abstract— **Poor software quality due to failure to organize development processes using SPI is visible in small software companies, although these companies are significant to industry and world economy; the challenge of quality needs much more attention. The development of different frameworks to sort the development process in SSC continues to leave a lot to be desired. This work leads us to the development of a comparison framework which will ultimately end with an adaptable SPA and SPI framework for SSC. From a SLR to identify the factors affecting SSC and mapping them to the specific processes in which they occur, we analyze the software development environment and identify SPI frameworks that have registered success in similar environments of SSC to help us adopt best practices from which comparison is made to generate requirements for an adaptable SPI framework within context for small companies.**

I. Introduction

Most of software development establishments face teething troubles in their projects due to deficiency of utilization of best practices and standards[1-3]. According to Khokhar et al [4] 50% of all application do not encounter trade goals while 40% of software development ventures do not deliver intended return on investment due to lack of best practices. Most software process assessment (SPA) models and standards pay attention primarily on medium and large companies; this obscures the helpfulness and proficiency of small software companies (SSC) perhaps because of their particular characteristics and limitations [5-7].

The entire software engineering SE industry appreciates the value of SSCs in terms of their link in producing cherished products and services to the industry. Although software process improvement (SPI) offers SSC prospects and trials[8], the bulk of software companies in industry fall within the SSC size category. It is essential to note that for SSC to offer customers with quality products that meet client needs within time and budget, the SSC must address daily trials, which involve probing software processes to ensuring production of quality products[5, 9]. This is a dreadful task to SSC probably due to time, resources and actually it ends up not being the case.

The apparent value of a simple practice in SPI is easily done using low cost practices, that makes SPI inexpensive for SSC with inadequate resources, facilitating them to achieve significant paybacks minus the utilization of a disproportionate quantity of resources[10]. This situation has stifled quality in SSC especially in Africa that SPI is unheard of. An organization that is interested in improving quality by undertaking SPI will require to conduct a quick and inexpensive procedure which is a nightmare for SSC, and this explains why software practitioners in such situations should figure out which SPI frameworks for the varied available framework is the most relevant. Sizeable achievement has been recorded by larger software companies leaving SSC behind hence the need to pay attention to the SSC. In this spirit, SPI has been adapted to fit in specific contexts[11, 12]. Accordingly, this work attempts to develop a roadmap for a SPI comparison framework (SPICOF) that compares frameworks used to solve trials in similar software development environment (SDE), so as to adopt improvement actions (IA) that will be applied to solve the trials in the SDE for which the framework is being developed. The outcome of this work will eventually be used to formulate requirements for an adoptable SPA framework for SSC based on the characteristic factors that affect SSC for example those from Africa.

The rest of this paper is organized as follows, Section II is related work, section III covers our approach, section IV covers what will makes our work different, section V covers our own analysis and chapter VI concludes the paper with what works will follow.

II. Related work

A. Comparing SPI frameworks

The Comparison of SPI frameworks is not common perhaps because of low uptake of SPI and maybe most of the frameworks that exist have been designed to suit specific SDE[13, 14]. This comparison of these frameworks is such a challenge because they are comprehensive and have particular description. In choosing SPI framework, practitioners choose frameworks as favorite based on subjective reasons. In cases of SSC we have seen frameworks developed also to suite specific SDE and cases of comparison are also limited to renowned frameworks like CMM, ISO 15504 and ISO 9000 perhaps this is explained by the fact that choosing a framework against another is

This work was not supported by any organization

©2019, PTI

complicated. Comparisons majorly for two different reasons;

First, authors of new frameworks compare what has already been developed with what is being authored to determine the differences and the similarities. In [1] Zarour highlights what he describes as authors point of view, in which authors of new frameworks compare with existent frameworks. For example in [6], Anacleto et al compares RAPID, SPINI, FAME, TOPS and MARES. As authors of MARES they seem to check the difference and similarity of MARES and the other 4 frameworks using requirements. In another example McCaffrey et al [15] compares their proposed assessment model for SSC to other lightweight assessment methods.

Secondly, organizations without SPI strategy or with little knowledge on SPI can prudently compare different SPI frameworks to learn about other frameworks so as to conduct self-assessment to evaluate capability levels. Zarour et al[1] and Halvorlsen et at [2] discuss organizations need to compare frameworks to facilitate choice of SPI frameworks. For instance in [13] Tingey makes a thorough comparison of CMM, ISO 15504 and the Malcom Balgride National Quality Award(MBA) to expedite companies capacity to take knowledgeable choices. We also see in [16] Paulk comparing ISO 9000 and CMM to help organization make a clear distinction on the underlying philosophies, although he also highlights the fact that they address common concerns of quality and process management (PRM).

The two scenarios have often led to the use of the four types of comparison methods, as discussed by Halvorlsen in his taxonomy [2]. In his classification, he highlights, needs mapping comparison method, framework mapping comparison method, characteristics comparison method, and bilateral comparison method. Authors [1, 2, 6, 13, 15-17] have compared SPI frameworks with the stated comparison methods argued as follows:

1) Needs Mapping Comparison Method

Needs mapping is a technique which contemplates organizational and environmental requirements when picking which SPI framework to embrace. Here are some examples as cited by [2]: Certification desires, for example of ISO 9000, is regularly enforced on subcontractors; Management may oblige that the preferred SPI framework must be incorporated in a total quality management methodology. In alternative occurrences, [1] uses this technique to compare TOPS, Micro-Evaluation (OWPL), MARES, SPM, RAPID, FAME, and EAP as different frameworks And similarly [6] compares MARES to RAPID, SPINI, TOPS, and FAME. This method is not a straight comparison between frameworks, in that requirements are considered to be of paramount significance and must be cautiously considered into since they are challenging and can utterly limit the select of framework. For the reason that evolution in organizational environment is perpetual, we see the requirements keep changing instigating variation from one organization to the other.

2) Framework Mapping Comparison Method

In [13] we appreciate that every single framework consists of a conventional statements or necessities allocating the content and emphasis of the framework as deliberated by [2]. Framework mapping is the procedure of constructing a map from statements or conceptions of one framework to those of the alternative. There are two divergent ways to do this: Mapping prevailing frameworks into the suggested framework being exclusively for the resolve of comparing, this delivers a mutual argument of reference from which the frameworks can be appraised, compared and contrasted; in the characteristics method the objective is to pronounce crucial traits of each SPI framework. Nevertheless, the purpose of mapping is to identify intersections and interactions between frameworks and produce a map of these statements or relationships. There can occur solid, weak or certainly not correlation as seen in Tingey's comparison of the 3 frameworks [13]. He, maps CMM, ISO9000 and the MBA on 2 stages, high and low reliant on the quantity of fact every one encompasses. This also decrees how the mapping outcomes can be revealed, for example in a matrix or some kind of graphic illustrations like Venn diagrams.

This is principally handy when an organization engages two or more distinct SPI frameworks, as analogous statements can be recognized and redundancy eliminated. The additional exertion required to employ an extra framework can be decreased. It is definitely a more low-level and detailed comparison method than characteristics. Because mapping goes into the specifics of each framework, it is not suitable for a broad impression. On the other hand, mapping into a base framework and supplementing with a quantitative analysis can point to overall focus and content. Framework mapping necessitates that some streamlining assumptions are put into thought. The outcomes are unavoidably prejudiced by these expectations as observed in [13].

3) Characteristics Comparison Method

Grounded on a far-reaching literature inspection, [2] and [1] compare various frameworks in a taxonomy by defining an extensive list of relevant characteristics. Each framework is then described in terms of these characteristics and the outcome is presented in a tabular format. This type of comparison is well suited for a general overview of the frameworks and it can be used as a foundation for other comparison methods. A lot of information can be concluded from a table listing the characteristics of a number of frameworks. However, the comparison is on a high level and details must be collected elsewhere. For this purpose one of the other comparison methods can be used.

In [1] we see five frameworks of RAPID, SPINI, FAME, TOPS and their new method MARES compared by Anacleto et al. the characteristics listed are objective, measurable and comparable. Although Halvorsen [2], also uses this framework in his taxonomy he argues that being objective is almost unmanageable in the firm sense, determinate is often a matter of human ruling, and comparable means that a practitioner will benefit from the comparison. He claims that a kind of dimension scale should

be used, e.g. minimal, ordinal or outright have diverse strong point and weak points, and numerous are used later. He parallels, six SPI frameworks; TQM, CMM v1.1, ISO 9000, ISO/IEC 15505(SPICE) GQM and SPIQ with a long list of characteristics

4) Bilateral Comparison Method

In this comparison method, we compare two frameworks textually, such as what we see when CMM and ISO 9000 actuality compared by Paulk in [16] then in another instance SPICE and ISO 9000 compared by El-Emam et al.in [17] The variance in the middle of this comparison technique and the others formerly designated is the textual natural surroundings. A bilateral comparison is frequently a rapid or description of outcomes from supplementary comparison techniques.

The bilateral comparison can carry on the argument of an outlook of one framework and refer to the other in terms of that. This is appropriate for people with thorough awareness of one framework, as they can easily get insight into other using familiar terms.

The extent of detail involved in a bilateral comparison can contrast extensively, depending on the purpose for which it is in black and white. Generally, the extent of fact is someplace in the middle of the characteristics and the mapping methodology.

III. Our Approach

A. Overview

Our focus is to have a SPICOF that compares frameworks that have been used in a similar SDE and also compare the challenges in software development. This is done to identify potential IA in the proposed framework from the action taken in solving the identified problems and resolving issues in the identified SDE. These potential IAs are scrutinized and mapped to the respective challenges for adoption of requirements and as basis for assessment in the proposed SPI framework being designed.

To put this into perspective we use data from the SLR conducted by [18] from which 69 frameworks for SPI used in SSC from 77 empirical studies over a period of thirty years, 21 frameworks are identified with significant frequencies of 80.2% with each having a frequency of at least 3 and classify the remaining 48 which appear once and twice as others with 19.8%. Frameworks CMMi and ISOIEC15504 stand out as represented in figure 1 with 15.8 % and 12 % respectively majorly because most of the other frameworks adopt certain principals from the two.

Table I shows process area and process extracted during the same SLR in [18] that identifies the 22 processes areas from 4 overall processes i.e. Project management (PM), Eng, PRM and Support, that are addressed using the SPI frameworks recognized in the empirical study, the process areas with the corresponding frequency of the factors affecting SSC showing which particular process they affect.

PM, ENG, Support SS, and PRM. PRM is most affected process in that order as illustrated in Fig 2.

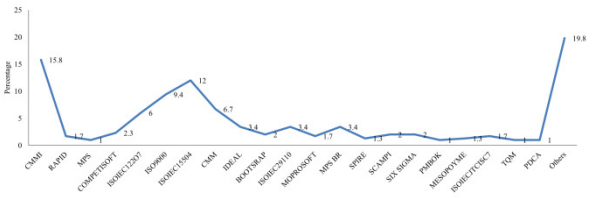


Fig 1 Graph representing SPI frameworks identified in the SLR frequencies in percentages

TABLE I
PROCESS AREAS, PROCESSES AND CORRESPONDING FREQUENCIES OF IDENTIFIED FACTORS AFFECTING SSC WITHIN THE PROCESSES

A	B	C	D
Project management	Project Planning and Organization	18	69
	Project Characteristics	1	
	Project Management	31	
	Subcontract management	3	
	Cost estimation	1	
	Risk Control and management	5	
	Tracking and oversight	3	
	Project Monitoring and Control.	7	
Engineering	Requirement gathering	1	49
	Requirements Management,	20	
	Requirements Assessment	1	
	Requirement development	3	
	Software Development	3	
	Verification and Validation	9	
	Configuration Management	12	
Process management	Measurement and analysis	2	13
	Tracking and oversight	3	
	Process Establishment	7	
	Change management	1	
Support	Process and Product Quality Assurance	29	31
	Problem Resolution	1	
	Management support & Commitment	1	

A- Process
B- Process Areas
C- Frequency of factors affecting SSC by process area
D- Frequency of factors affecting SSC by process

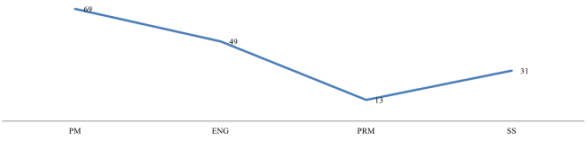


Fig 2. Percentile representation of factors affecting SSC against process areas

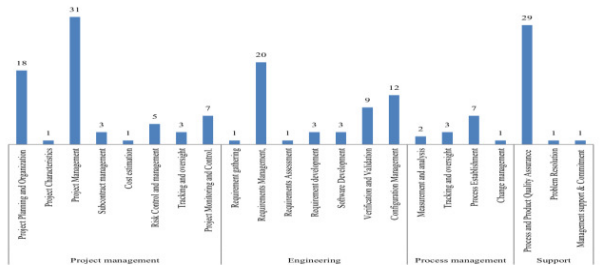


Fig 3. Graph representing frequencies of factors affecting SSC in specific processes

The 69 frameworks used for SSC identified from the SLR give us a picture of how much work has been done despite the continued slow pace of improvement in quality of software especially in SSC.

B. Analysis of different frameworks and how they are used in different environment

Based on the analysis of the SPI frameworks for SSC we single out 7 based their characteristics and the development environment from which they operate i.e. Moprosoft, Mares, iSPA, RAPID, SPINI, MPS-BR and MECA see table III. We study the activities involved the improvement process of each of the identified SPI frameworks and discuss them as follows;

1) MoProSoft

This framework was established in Mexico together with the UNAM, the Mexican Association for the Quality in Software Engineering (AMCIS) and the Ministry of Economy[20], The drive of the manuscript was to present a Software Industry Processes Model (MoProSoft) in Mexico supporting the regularization of its procedure through the integration of best practices in software management and Eng.

Adopting the model was meant to permit nurturing the capability of organizations to offer high quality services and reach international level of competitiveness.

It has 6 processes; business management; PRM; PM; resource management; administration of definite assignments and Software development centered on the ISO/IEC/12207, CMM, ISO9001[19]

2) The MARES Process Assessment Model

Brazilian software industry has grown at a fast rate yet playing a pivotal role in the economy. However, the SSC were face quality and productivity problems caused by organizational and administrative deficiencies, which affected their competitiveness.

MARES as a methodology for SPA in SSC was designed to support software process considering specific characteristics and limitations[7, 20]. SSC in Brazil found it particularly difficult to run assessments in conformance with international standards or models, such as, ISO/IEC 15504, CMMI for software or ISO 9001 (including 9000-3) due the major emphasis of these models and standards on medium or large companies. Nevertheless, characteristics and confines typical for SSC in conformance with ISO/IEC 15504 by integrating context –process model in order to support the selection of relevant processes and process-risk model to support the identification of potential risks and improvement suggestions [21].

The MARES process assessment framework was also enhanced by the description of a context-process relationship model, which mimics the association between specific characteristics known difficulties and business objectives[6]. It was assembled by researchers from the UNIVALI University in collaboration with CenPra research center in Brazil[22].

MARES is divided into five core parts which are planning, contextualization, execution, monitoring and control and post-mortem[6]. Methodology for process assessment in SSC entail a model which is based on the exemplar model of part 5 of ISO/IEC 15504, [20] an assessment process that encounters the requirements of the assessment process defined in ISO/IEC 15504-2, an assessor accreditation method which outlines procedures and an assessment monitoring method which enables the constant monitoring of assessment methodology. [7, 22]

3) iSPA Integrated Software Process Assessment

The Malaysian software industry was facing difficulties in their software projects due to lack of putting into practice the best practices and standards. According [23] 50% of all application are unsuccessful and did not achieve business goals whereas 40% of software projects failed to deliver intended return on investment due to lack of best practices. Among the causes of lack of best practices was level of awareness and understanding amongst software development companies. There were also issues of costs involved in the direction of the execution of standards or benchmark requirements which had become a significant deterrent factors [24]. This gave rise to the development of the SPA framework iSPI as a starting point from the ISO/IEC 15504 (Information Technology – Process Assessment) as an all-inclusive standards or SPA [25]. ISO/IEC 15504 is seen as an internationally recognized standard by ISO organization, and also embraced as a Malaysian standard. The framework has 3 layers to the system which are presentation, application and data layers with 6 processes Project/Process profiling gap analysis; Strength/Weakness generator; Improvement plan; Customization of PRM and Reporting[25].

4) RAPID (Rapid Assessment or Process Improvement for software Development)

This is a light weight ISO/IEC 15504 or SPICE complaint assessment methodology with a range of 8 processes[26]. Requirement elicitation, software development, configuration management, quality assurance, problem resolution, PM, risk management, and process establishment. This methodology has been preferred for evaluation to its semblance with SPICE[27] probably because of its addition of threat management process within its activities.

TABLE II.
SELECTED SPI FRAMEWORKS AND CORRESPONDING PROCESSES ADDRESSED DURING ASSESSMENT AND IMPROVEMENT

software process areas	Software processes	CMMI	RAPID	iSPA	MARES	SPINI	MOPROSOFT	MPS BR	MECA
Project management	Project Planning and Organization	x		x	x	No specific number of processes		x	No specific number of processes
	Project Characteristics	x					x	x	
	Project Management	x	x				x	x	
	Subcontract management	x						x	
	Cost estimation	x						x	
	Risk Control and management	x	x	x				X	
	Tracking and oversight	x						X	
	Project Monitoring and Control.	x		x	x		x	X	
Engineering	Requirement gathering	x	x					X	
	Requirements Management,	x		x				X	
	Requirements Assessment	x			x			X	
	Requirement development	x						X	
	Software Development	x	x		x		x	x	
	Verification and Validation	x						x	
	Configuration Management	x	x					x	
Process management	Measurement and analysis	x		x				x	
	Tracking and oversight	x			x			x	
	Process Establishment	x	x				x	x	
	Change management	x						x	
Support	Process and Product Quality Assurance	x	x					x	
	Problem Resolution	x	x	x				x	
	Management support & Commitment	x					x	x	

The key aim of RAPID is to pinpoint the strong points of companies and also make certain threats and improvement prospects[7]. The Rapid assessment encompasses 210 questions which will be appraised concentrating on risk management practice and analysis of the consequence of the process.

The RAPID assessment was constructed following the findings from a technical report, Much of the assessment is done through an interview it places its significant importance on the competence of the person doing the assessment, the assessor assumes dual roles which are team leader and support assessor, in each of the roles the assessor will have a specific task to carry out[28] RAPID was established by the Australian Software Quality Institute [29].

5) SPINI (An approach for SPI Initiation)

SPINI is an ISO/IEC 15504 or SPICE compatibility assessment method it was recognized as part of the SataSPIN project it was introduced in 1998 in the Satakunta region Western Finland[7]. This initiative was established to advantage software companies most of which ware part of SMEs to grow their operations with known standard frameworks. SPINI encompasses 3 agendas; First of all, the organization desires to appreciate the abilities of SPI in achieving its business objectives. Secondly, an assessment of the current software processes desires to be complete and lastly the SPI actions require be planned and supported. The framework comprises of events that are assembled from peripheral advisors; points of view steps; understanding priorities, carrying out assessment and supporting SPI. [14] The method articulates the important actions that were established as beneficial in starting up SPI using external support [7].

6) MPS-BR

MPS.BR Program is a countrywide mobilization drive that was envisioned for the improvement of software process in Brazil which created the MPS Model.

The MPS.BR program acronym is in Portuguese meaning “Melhoria de Processo do Software Brasileiro” was created in late 2003 by an Association for Promoting the Brazilian Software Excellence (SOFTEX). This association is composed of organizations in some cities of Brazilian states. SOFTEX has about 1,600 conglomerated companies and 70% are SSC.

The MPS-BR model constitutes of three key components: MPS Reference Model (MR-MPS); MPS Assessment Method (MA-MPS); and MPS Business Model (MN-MPS). MR-MPS and MA-MPS are ISO/IEC 12207 and ISO/IEC 15504, CMMI-DEV compatible, based on SE best practices, and in harmony with the Brazilian companies reality.

The three key components consist of 23 processes organization, innovation and deployment; Causal analysis and resolution; Organizational process performance; Qualitative PM Risk management; Decision analysis and resolution; Requirement development; Technical solution; Validation; Verification; Software integration; Software installation; Product release; Training; Process establishment; Process assessment and improvement; Tailoring process for PM; Configuration management; QA; Acquisition; Measurement; PM and Requirement management.

7) MECA

In Pakistan most large, medium and SSC had challenges in establishing quality improvement programmes. Contrary to others, Khokhar et al [4] perceive large organizations as distinct to SSC finding them generally complex to implement and adjust to quality improvement strategies. This proved complex since most quality models like CMMI and SPICE primarily addressed the requirements of large companies only. The MECA model offers a continuous monitoring approach for software processes and provides basis to commence improvement.

MECA is motivated by the Plan, Do, Check Act (PDCA) model. The PDCA was established by Walter Shewhart, it is responsible for the basic viewpoint of a controlled, cyclic approach to continuous improvement. MECA aids companies which continuously struggle to attain CMMI.

Thereafter the output of the SPI framework designed will also be used to compare the existent outcome of adopted SPI frameworks. We use two different SPI comparison methods, the characteristics method and the needs mapping comparison method. Fig.5 is a UML notation of a proposed dynamic SPICOF which has 3 main components being Analysis, Mapping IA and requirements and has 8 processes. The requirements are then used for the design and development of a SPI framework which is external to this work to see fig 6.

C. The proposed SPICOF

1) Analysis

In the first component denoted as 1 the proposed framework has 2 steps one is to analysis of SDE to elicit the challenges and the environmental context in which software is being developed. Then the second step identifies software frameworks that have been used in similar SDE to study the IA from which the existent frameworks being compared. These are denoted with numbers 1,2,3,…n in fig. 4. This serves to elicit the potential IA on the challenges that were addressed and in the SDE. The number of frameworks to select for comparison depends on existent frameworks chosen, preferably a choice of frameworks that addresses at list one of the challenges and SDE issues elicited. The output of step one and two in the first component of the proposed SPICOF are then tabulated in a matrix as input in the second component of the proposed SPICOF.

2) Mapping Improvement action to challenges

This component of the proposed SPICOF has 2 processes, 3 and 4, the first process identifies which potential IAs should be taken, this IA is then mapped to a corresponding challenge and SDE issues listed in the matrix. This potential IA in the matrix is entered so as to give a basis from which choice will be made for the best. In same step the different potential IAs identified in the matrix are compared to choose the best option that can solve the challenge and the SDE issues identified.

Challenges and SDE issues that have not been mapped to a potential IA are put aside in the 4th process for other challenges that have not been mapped to IAs in the matrix to be identified as a candidate challenges for further scrutiny whose IA are sought from other best practices. The mapping is done at low level so as to scrutinise all siginificant detail required to solve particular challenge.

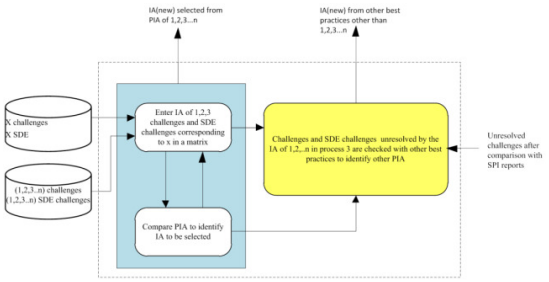


Fig 4. Mapping improvement action to challenges

3) Requirements for SPI

Component 3 has three processes, 5 through 7 with iteration between step 7 back to the 4th process in the second component and then to the 5th, 6th and 7th in the third component. In step 5 candidate improvement actions are identified as requirements using the characteristics comparison method the requirements are scrutinized for presentation as measureable, comparable and extremely objective characteristics being input for construction of the SPI framework. An analysis is performed and results are mapped to output of step 1 and 2 for consistency and traceability. In this step again uses characteristics comparison method which is based on extensive literature survey in steps 1 and 2. The requirements that are defined in terms of the chosen potential IAs and input for step 6.

The actual construction of an adaptable SPI framework is then done using data from this stage SPA and SPI is done outside this scope see fig 6. All chosen IAs from different frameworks are put together to address the challenges identified, environmental issues elicited and results of the assessment from the SPA conducted.

In proses 7 the final comparison of the report of the SPI results highlighting the impact is compared with impacts of IAs conducted from the candidate actions adopted. After this assessment the gaps identified are iterated through step 4, 5, 6, and 7 until marginal gaps that can be negligible are identified then we stop here.

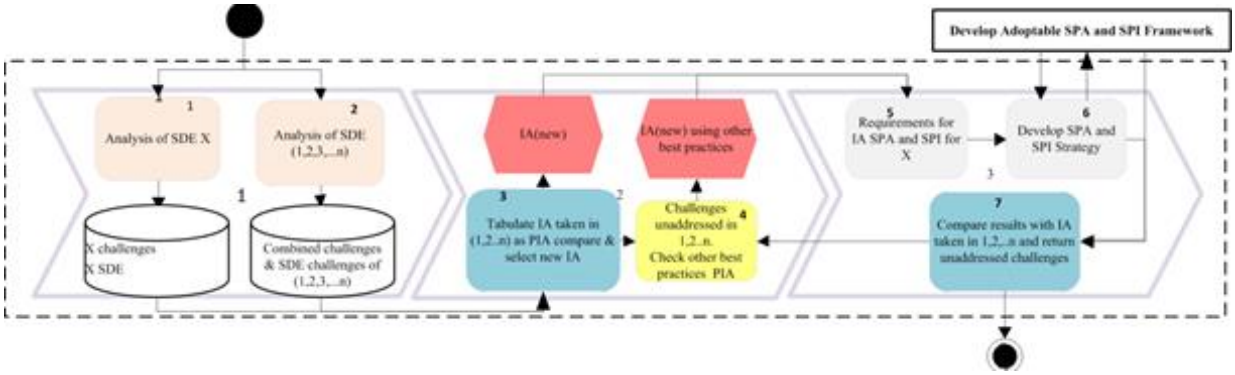


Fig 5. Dynamic SPI comparison framework for generating requirements for an adoptable SPI framework

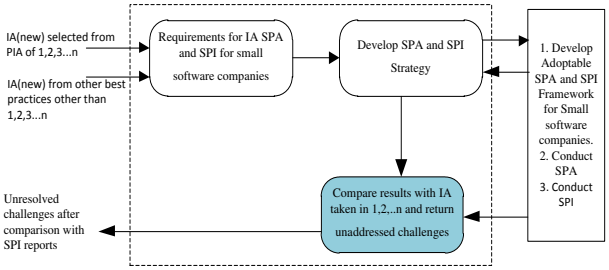


Fig 6. Requirements Improvement actions

IV. WHAT MAKES OUR FRAMEWORK DIFFERENT

A. *Overview*

SPA and SPI frameworks are significant in software process modeling and improvement to the extent that where software development is being practiced without SPI has a lot to be desired. This framework paves way for simplifying SPA and API for SSC. Unlike other frameworks this proposed framework compares SPI frameworks for the purpose of benchmarking to design and develop an adaptable SPI framework. In table II we compare the characteristics of our framework with the frameworks of [1, 2, 6].

The proposed framework has a twofold comparison; this makes it unique from the other SPICOFs that look at comparison only in perspective of what has been done. We compare for assessment before process improvement and for assessment of our SPI framework after application.

B. *Assessment of proposed framework*

Once characteristics of the company and the SDE in which they operate are identified, IA are identified and improvement is conducted then an evaluation based on test of safeguards in operational planning, the flexibility, informal management styles and risk management to check the changes in the safeguards still affect the processes. And then back to the loop.

V. OUR OWN ANALYSIS

A dilemma came into SE around 1965-1985 when failure of projects was increasing and the main cause of this crisis was the overall complexity of the software process, with projects over shooting budgets; and schedule; Low quality software; Requirement are not met; generally, projects were troublesome [30]. Accordingly, different solutions came up including SPI with frameworks and models like CMMI and SPICE.

Frameworks are reusable design for all or parts of development that is represented by a set of abstract processes, describing how software development is broken down into a set of interacting processes and the interaction with in the sub processes. If only SSC could effectively use frameworks, There is a promising avenue in this for achieving best practices in software development particularly by reducing the cost and ultimately improving the quality of software. Frameworks have played a pivotal role in software process modeling, evaluation and improvement; they have

TABLE III.
A COMPARISON OF CHARACTERISTICS OF PROPOSED COMPARISON FRAMEWORK TO OTHER COMPARISON FRAMEWORK

Zarour et al. Characteristics[1]	Halvorsen Characteristics[2]	Anacleto et al. Characteristics[6]	Proposed Characteristics
-Total of evaluated processes -Assessed processes. -Number of processes to be enhanced -Assessment period	-Geographic origin/spread -Scientific origin -Development/firmness -reputation -Analysis procedures	-Budget -Guidance for process choice -Support for identification of threat and perfection proposals -Requirement for precise SE awareness from the company representative -Tool provision -Public handiness	-Numbers of challenges being addressed. -Similarity of SDE in terms of skill, customer awareness, PM practices. -How challenges addressed vs adopted of tools. -Impact on quality of software. -Timeliness to market, team size vs project completion time

also enabled developers to rapidly produce new applications. However, the absences of an adaptable framework for different contexts in SSC make it difficult to improve the processes. An understanding of the characteristics of SSC will fill the gap that has made comparison of different SDE an attainable.

Most organizations fail to adopt such frameworks due to the costs involved, [31], and time factor and yet the other SPI frameworks continue to adopt the already marked as complicated frameworks. Although performing the CMMI based improvement is difficult and expensive for SSC [14]. Small companies often fail to afford the cost overrun, which is necessary for the popular frameworks.

In Africa for example, the case is different although the SSC face the same challenges of lack of implementation of best practices, lack of adopted standards and generally disorganized software processes; there is also no evidence in literature, of attempts to localize the existing frameworks for SPI[32-36].

Countries like Brazil, Mexico, Pakistan, Australia and Malaysia have developed local standards like MPS-BR, Moprosoft, iSPA, MARES, SPINI and MECA [9, 37, 38]. Because the existing international frameworks like CMMI and ISO 15505 have not registered substantial successes with SSC due to different reasons, all attempts to improve software process have not been ceased.

The success of a better software industry in places like Africa will be dependent on how much effort can be put to improve the status quo[36]. It is therefore high time since the industry is growing rapidly that we began to assess the software processes in Africa, in order to improve on quality of software.

Software process is a set of related activities that split software development efforts into separate stages to improve design, product management, and PM; it leads to the production of software. These activities may involve the development of the software from the basic idea, or transforming existing software. Processes are significant for providing avenue for efficiency, effectiveness, and reduce waste in an organization. Ineffective processes cause frustration, delays, and increased costs, however they can be improved.

SPI frameworks in SSC face numerous challenges, most of all is the ability to demonstrate the expected business results as suggested by Wangenheim et al. [39] and Cater-Steel et al. [40]. These authors suggest the reduction in process assessment expenses and the time essential to make the SPI benefits noticeable. Accordingly Staples et al. [41], also proposes that SPI approaches dedicated to SSC are better cheaper and within a reasonable time for SSC.

VI. CONCLUSION AND FUTURE WORK

Our findings indicate that contextualization of SPI require each specific context to be treated differently and hence the need for an adaptable framework. We have noted that the SDE for SSC significantly differ and the difficulty in one place may also occur differently and this calls for tailoring the solution differently.

We have observed that the application of the CMMI, SPICE and other standard frameworks and there adoption for the purpose of design of SPI frameworks may not necessarily take us to successful results for SSC. The software development processes in SSC and the factors that affect them do not occur in the same pattern project management process area may have less issues compared to PRM process area in two software different environments owing to different reasons.

These observations make the road map for the development of a SPICOF clearer giving us good insights into the development of the adaptable SPA and SPI framework for SSC which is context independent. Therefore the framework being developed will draw lessons from the above observations and other findings next studies to identify characteristics of SSC. In the near future we plan to make comparison of different SDE in relation to the SSC characteristics and how much it affects quality before we validate the SPICOF being developed for SPI in SSC.

REFERENCES

[1] M. Zarour, J.-M. Desharnais, and A. Abran, "A framework to compare SPAt methods dedicated to small and very small organizations," *International Conference on Software Quality-ICSQ*, 2007, pp. 16-17.

[2] C. P. Halvorsen and R. Conradi, "A taxonomy to compare SPI frameworks," *European Workshop on Software Process Technology*, 2001, pp. 217-235. http://dx.doi.org/10.1007/3-540-45752-6_17

[3] A. M. AL-Ashmori, B. B. Rad, and I. Suhaimi, "SPI frameworks as alternative of CMMI for SMEs: a literature review," *Journal of SE,* vol. 11, pp. 123-133, 2017. http://dx.doi.org/10.3923/jse.2017.123.133

[4] M. N. Khokhar, A. Mansoor, K. Muhammad Nadeem, R. Shafiq Ur, and A. Rauf, "MECA: SPI for small organizations," *International Conference on Information and Emerging Technologies*, 2010, pp. 1-6. http://dx.doi.org/10.1109/iciet.2010.5625678

[5] A. A. Khan, J. Keung, M. Niazi, S. Hussain, and M. Shameem, "GSEPIM: A roadmap for SPA and SPI in the domain of global software development," *Journal of software: Evolution and Process,* vol. 31, p. e1988, 2019. http://dx.doi.org/10.1002/smr.1988

[6] R. Anacleto, C. G. Von Wangenheim, C. F. Salviano, R. Savi, S. Brasil, and C. S. Brasil, "A method for process assessment in small software companies," 2004.

[7] C. G. von Wangenheim, A. Anacleto, and C. F. Salviano, "Mares-a methodology for software process assessment in small software companies," Technical Report LPQS0012004.

[8] X. Larrucea, R. V. O'Connor, R. Colomo-Palacios, and C. Y. Laporte, "Software process improvement in very small organizations," *IEEE Software,* vol. 33, pp. 85-89, 2016. http://dx.doi.org/ 10.1109/ms.2016.42

[9] T. Tahir, G. Rasool, and M. Noman, "A systematic mapping study on software measurement programs in SMEs," *e-Informatica Software Engineering Journal,* vol. 12, 2018. http://dx.doi.org/ 10.1109/icacs.2018.8333491

[10] C. Y. Laporte, J.-M. Desharnais, M. M. Abouelfattah, J.-C. Bamba, A. Renault, and N. Habra, "Initiating SPI in Small Enterprises: Experiments with CETIC's Micro-Evaluation Framework," *International Conference on Software Development (SWDCREK 2005) Reykjavik, Iceland*, 2005.

[11] E. Mnkandla, "About software engineering frameworks and methodologies," in AFRICON 2009, 2009, pp. 1-5. http://dx.doi.org/10.1109/afrcon.2009.5308117

[12] J. Nakatumba-Nabende, B. Kanagwa, R. Hebig, R. Heldal, and E. Knauss, "Hybrid Software and Systems Development in Practice: Perspectives from Sweden and Uganda," International Conference on Product-Focused SPI, 2017, pp. 413-419. http://dx.doi.org/10.1007/978-3-319-69926-4_30

[13] M. O. Tingey, Comparing ISO 9000, Malcolm Balbridge, and the SEI CMM for Software: A Reference and Selection Guide: Prentice Hall PTR, 1996.

[14] C. Y. Laporte, S. Alexandre, and R. V. O'Connor, "A software engineering lifecycle standard for very small enterprises," European Conference on SPI, 2008, pp. 129-141. http://dx.doi.org/10.1007/978-3-540-85936-9_12

[15] F. McCaffery, D. McFall, and F. G. Wilkie, "Improving the express process appraisal method," International Conference on Product Focused SPI, 2005, pp. 286-298. http://dx.doi.org/10.1007/11497455_24

[16] M. C. Paulk, "How ISO 9001 compares with the CMM," IEEE software, vol. 12, pp. 74-83, 1995. http://dx.doi.org/10.1109/52.363163

[17] K. E. Emam, W. Melo, and J.-N. Drouin, SPICE: The theory and practice of SPI and capability determination: IEEE Computer Society Press, 1997.

[18] M. Tuape and Y. Ayalew "Factors affecting development process in small software companies," IEEE/ACM Symposium on Software Engineering in Africa (SEiA) 2019. pp. 16-25 http://dx.doi.org/10.1109/SEiA.2019.00011

[19] M. E. Morales-Trujillo, H. Oktaba, T. Ventura, and R. Torres, "From MoProSoft Level 2 to ISO/IEC 29110 Basic Profile: Bridging the Gap," Clei Electronic Journal, vol. 16, 2013. http://dx.doi.org/10.19153/cleiej.16.1.2

[20] V. Ribaud and P. Saliou, "Process assessment issues of the ISO/IEC 29110 emerging standard," in Proceedings of the 11th International Conference on Product Focused Software, 2010, pp. 24-27. http://dx.doi.org/10.1145/1961258.1961264

[21] C. G. von Wangenheim, A. Anacleto, and C. F. Salviano, "Helping small companies assess software processes," IEEE software, vol. 23, pp. 91-98, 2006. https://doi.org/10.1109/ms.2006.13

[22] F. J. Pino, C. Pardo, F. García, and M. Piattini, "Assessment methodology for SPI in small organizations," Information and Software Technology, vol. 52, pp. 1044-1061, 2010. https://doi.org/10.1016/j.infsof.2010.04.004

[23] R. Z. R. M. Ali and S. Ibrahim, "An application tool to support the implementation of integrated SPIt for Malaysia's SME," Malaysian Conference in Software Engineering, 2011, pp. 177-182. https://doi.org/10.1109/mysec.2011.6140665

[24] R. Ali and S. Ibrahim, "An Integrated SPA for Malaysia's SME," Postgraduate Annual Research Seminar (PARS) in UTM Skudai, 2009.

[25] R. Zurina, R. M. Ali, and S. Ibrahim, "An iSPA model evaluation based on critical success factors and selected criteria to support Malaysia's SME environment," 2nd International Conference on Software Engineering and Data Mining, 2010, pp. 225-230.

[26] A. Cater-Steel, M. Toleman, and T. Rout, "Process improvement for small firms: An evaluation of the RAPID assessment-based method," Information and Software Technology, vol. 48, pp. 323-334, 2006. https://doi.org/10.1016/j.infsof.2005.09.012

[27] T. P. Rout, "ISO/IEC 15504 and Spice," Encyclopedia of Software Engineering, 2002. https://doi.org/10.1002/0471028959.sof171

[28] T. Rout, "Evolving SPICE–the Future for ISO 15504," 1st International Conference on SPI and Capability Determination, Limmerick, Ireland, 2000, pp. 1-8.

[29] H. Oktaba and M. Piattini, "SPI for Small and Medium Enterprises: Techniques and Case Studies," 2008. https://doi.org/10.4018/978-1-59904-906-9

[30] E. Klotins, M. Unterkalmsteiner, and T. Gorschek, "Software engineering in start-up companies: An analysis of 88 experience reports," Empirical Software Engineering, vol. 24, pp. 68-102, 2019. https://doi.org/10.1007/s10664-018-9620-y

[31] T. Dybå, "Enabling SPIt: an investigation of the importance of organizational issues," Empirical Software Engineering, vol. 7, pp. 387-390, 2002. https://doi.org/10.1023/a:1020535725648

[32] C. Ferreira and J. Cohen, "Agile systems development and stakeholder satisfaction: a South African empirical study," South African Institute of Computer Scientists and Information Technologists on IT research in developing countries: riding the wave of technology, 2008, pp. 48-55. https://doi.org/10.1145/1456659.1456666

[33] B. De Wet and J. Visser, "An evaluation of software project risk management in South Africa," South African Journal of Industrial Engineering, vol. 24, pp. 14-29, 2013. https://doi.org/10.7166/24-1-497

[34] E. Kabaale, G. Mayoka Kituyi, and I. Mbarika, "Requirements engineering process improvement challenges faced by software SMEs in Uganda," 2014. https://doi.org/10.5120/15349-3690

[35] J. K. Balikuddembe and M. Tuape, "An Ambiguity Minimization Technique during Requirements Elicitation Phase," International Conference on Computational Science and Computational Intelligence (CSCI), 2017, pp. 945-950 https://doi.org/10.1109/csci.2017.164.

[36] Y. Ayalew and K. Motlhala, "Software Process Practices in Small Software Companies in Botswana," 14th International Conference on Computational Science and Its Applications, 2014, pp. 49-57. https://doi.org/10.1109/iccsa.2014.20

[37] A. A. Khan, J. Keung, S. Hussain, M. Niazi, and M. M. I. Tamimy, "Understanding SPIt in global software development: a theoretical framework of human factors," ACM SIGAPP Applied Computing Review, vol. 17, pp. 5-15, 2017. https://doi.org/10.1145/3131080.3131081

[38] V. Garousi, M. Felderer, and T. Hacaloğlu, "What we know about software test maturity and test process improvement," IEEE Software, vol. 35, pp. 84-92, 2018. https://doi.org/10.1109/ms.2017.4541043

[39] C. G. v. Wangenheim, A. Anacleto, and C. F. Salviano, "Helping small companies assess software processes," IEEE Software, vol. 23, pp. 91-98, 2006. https://doi.org/10.1109/ms.2006.13

[40] A. P. Cater-Steel, "Process improvement in four small software companies," Australian Software Engineering Conference, 2001, pp. 262-272. https://doi.org/10.1109/aswec.2001.948520

[41] M. Staples and M. Niazi, "Two case studies on small enterprise motivation and readiness for CMMI," 11th International Conference on Product Focused Software, Limerick, Ireland, 2010. https://doi.org/10.1145/1961258.1961274

Author Index

Albuquerque, Robson de O. 81
Alencastro, João F. de 81
Almeida, Thiago S. 3
Aszalós, László 13, 21
Ayalew, Yirsaw 133

Bakó, Mária 21
Banach, Marzena 67, 73
Benhammadi, Farid 109
Beširević, Admir 125
Bey, Kadda Beghdad 109
Bouznad, Sofiane 109

Delalić, Sead 125
De Sousa, Jr, Rafael T. 81
Długosz, Rafał 67, 73
Drahoš, Peter 49
Dutra, Bruno V. 81

Elgammal, Amal 93
Elzanfaly, Doaa 93

Filho, Francisco L. de Caldas 81

Gonçalves, Daniel G. V. 81
Gonsorčík, Jozef 119

Haffner, Oto 49
Hetmaniok, Edyta 41
Hodžić, Kerim 125

Jarčuška, Peter 119
Júnior, Methanias Colaço 3
Junosza-Szaniawski, Konstanty 29

Kfouri, Guilherme de O. 81
Kubiak, Katarzyna 67, 73
Kučera, Erik 49

Leskovský, Roman 49

Majerník, Jaroslav 119
Martins, Lucas M. C. e 81
Matišák, Jakub 49

Nacer, Hassina 109
Nagy, Dávid 13
Nascimento, Anderson Clayton Alves 57
Nogalski, Dariusz 29

Ordonez, Edward David Moreno 57

Pleszczyński, Mariusz 41
Praciano, Bruno J. G. 81

Riad, Maha 93
Ribeiro, Admilson de Ribamar Lima 57
Różański, Michał 41

Silva, Victor Matheus da 3
Słota, Damian 41
Szczygieł, Marcin 41

Tuape, Micheal 133
Tucaković, Zlatan 125

Vasconcelos, Nathanael Oliveira 3

Weichbroth, Paweł 101
Wituła, Roman 41

Žunić, Emir 125