Sotiris Batsakis
Heinrich C. Mayr
Vitaliy Yakovyna
Mykola Nikitchenko
Grygoriy Zholtkevych
Vyacheslav Kharchenko
Hennadiy Kravtsov
Vitaliy Kobets
Vladimir Peschanenko
Vadim Ermolayev
Yuriy Bobalo
Aleksander Spivakovsky
(Eds.)

ICT in Education, Research and Industrial Applications: Integration, Harmonization and Knowledge Transfer

Proceedings of the 11th International Conference, ICTERI 2015

Lviv, Ukraine
May, 2015

Batsakis, S., Mayr, H. C., Yakovyna, V., Nikitchenko, M., Zholtkevych, G., Kharchenko, V., Kravtsov, H., Kobets, V., Peschanenko, V., Ermolayev, V., Bobalo, Yu. and Spivakovsky, A., (Eds.): ICT in Education, Research and Industrial Applications: Integration, Harmonization and Knowledge Transfer. Proc. 11th Int. Conf. ICTERI 2015, Lviv, Ukraine, May 14-16, 2015, CEUR-WS.org, online

This volume represents the proceedings of the 11th International Conference on ICT in Education, Research, and Industrial Applications, held in Lviv, Ukraine, in May 2015. It comprises 45 contributed papers that were carefully peer reviewed (3-4 reviews per paper) and selected from 119 submissions.

The volume opens with the abstracts of the keynote talks and tutorial. The rest of the collection is organized in 2 parts. Part I contains the contributions to the main ICTERI conference, structured in four topical sections: (1) Teaching ICT and Using ICT in Education; (2) Model-Based Software System Development; (3) Machine Intelligence, Knowledge Engineering and Management for ICT; and (4) ICT in Industrial Applications. Part II comprises the contributions of the four workshops co-located with ICTERI 2015, namely: the International Workshop on Information Technologies in Economic Research (ITER 2015); the International Workshop on Methods and Resources of Distance Learning (MRDL 2015); the International Workshop on Algebraic, Logical, and Algorithmic Methods of System Modeling, Specification and Verification (SMSV 2015); and the International Workshop on Theory of Reliability for Modern Information Technologies (TheRMIT 2015).

Preface

ICTERI, the *International Conference on Information and Communication Technologies in Education, Research, and Industrial Applications: Integration, Harmonization, and Knowledge Transfer*, has become a considerable and stable international ICT conference. It is a real pleasure for all ICTERI players, that in contrast to 2014, the 11th edition could bring scholars and expert representatives physically together again for exchanging and discussing new ideas and findings, and for networking across all political borders. This is all the more pleasing as, despite of all current challenges, the Ukrainian ICT community proves its vigor and global integration.

We gladly present you the proceedings of ICTERI 2015, which was held in Lviv, Ukraine, on May 14-16, 2015. The conference scope was determined by the cornerstones of *ICT Infrastructures and Techniques*, *Knowledge Based Systems*, *Academia/Industry ICT Cooperation*, and *ICT in education*. Special emphasis was given to real world applications of ICT solutions. Therefore, the contributions had to describe original, not previously published work, and to demonstrate how and to what purpose and extent the proposed solutions are applied or transferred into use.

For the main conference, 42 full papers were submitted and evaluated by at least three peers per paper. Finally, 16 have been selected and accepted after revision in accordance with the reviewers comments. This corresponds to an acceptance rate of 38%. The program was rounded off with the two outstanding keynote talks on *Rigorous Semantics and Refinement for Business Processes* by Klaus-Dieter Schewe and on *Smart Learning Environments: a Shift of Paradigm* by David Esteban. The tutorial on *Systematic Business Process Modeling in a Nutshell* by Heinrich C. Mayr complemented the program, in particular regarding the emphasis on the synergy of education and industrial applications.

ICTERI 2015 continued the tradition of hosting co-located events, this year by offering four workshops:

- 4th Int. Workshop on Information technologies in economic research (ITER 2015)
- 3rd Int. Workshop on Methods and Resources for Distance Learning (MRDL 2015)
- 4th Int. Workshop on Algebraic, Logical, and Algorithmic Methods of System Modeling, Specification and Verification (SMSV 2015)
- Int. Workshop on Theory of Reliability for Modern Information Technologies (TheRMIT 2015)

In total, these workshops attracted 77 submissions, from which 29 were selected by the particular program committees. This again led to an acceptance rate of 38%.

Clearly, the conference would not have been possible without the engaged support of many people including the authors, members of our Program Committee, workshop organizers and their program committees, local organizers, and, last but not least, generous donators. We express our special thanks to all of them.

May, 2015

Sotiris Batsakis, Heinrich C. Mayr, Vitaliy Yakovyna, Mykola Nikitchenko, Gryzoriy Zholtkevych, Vyacheslav Kharchenko, Hennadiy Kravtsov, Vitaliy Kobets, Vladimir Perchanenko, Vadim Ermolayev, Yuriy Bobalo, Aleksander Spivakovsky

Committees

General Chairs

Yuriy Bobalo, Lviv Polytechnic National University, Ukraine
Aleksander Spivakovsky, Kherson State University, Ukraine

Steering Committee

Vadim Ermolayev, Zaporizhzhya National University, Ukraine
Heinrich C. Mayr, Alpen-Adria-Universät Klagenfurt, Austria
Mykola Nikitchenko, Taras Shevchenko National University of Kyiv, Ukraine
Aleksander Spivakovsky, Kherson State University, Ukraine
Mikhail Zavileysky, DataArt, Russian Federation
Grygoriy Zholtkevych, V.N.Karazin Kharkiv National University, Ukraine

Program Chairs

Sotiris Batsakis, University of Huddersfield, UK
Heinrich C. Mayr, Alpen-Adria-Universität Klagenfurt, Klagenfurt, Austria
Vitaliy Yakovyna, Lviv Polytechnic National University, Ukraine

Workshop Chairs

Mykola Nikitchenko, Taras Shevchenko National University of Kyiv, Ukraine
Grygoriy Zholtkevych, V.N.Karazin Kharkiv National University, Ukraine

Tutorial Chair

Vadim Ermolayev, Zaporizhzhya National University, Ukraine

IT Talks Chairs

Aleksander Spivakovsky, Kherson State University, Ukraine
Mikhail Zavileysky, DataArt, Russian Federation

Local Organization Chair

Dmytro Fedasyuk, Lviv Polytechnic National University, Ukraine

Publicity Chair

Nataliya Kushnir, Kherson State University, Ukraine

Web Chair

Eugene Alferov, Kherson State University, Ukraine

Program Committees

MAIN ICTERI 2015 Conference

Program Committee

Jan Aidemark, Linneaus University, Sweden
Eugene Alferov, Kherson State University, Ukraine
Costin Badica, University of Craiova, Romania
Nick Bassiliades, Aristotle University of Thessaloniki, Greece
Sotiris Batsakis, University of Huddersfield, UK
Lukas Chrpa, University of Huddersfield, UK
Michael Cochez, University of Jyväskylä, Finland
Anatoliy Doroshenko, National University of Technology "Kyiv Polytechnic Institute", Ukraine
Vadim Ermolayev, Zaporizhzhya National University, Ukraine
David Esteban, TECHFORCE, Spain
Wolfgang Faber, University of Huddersfield, UK
Anna Fensel, STI, University of Innsbruck, Austria
Vladimir Gorodetsky, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Science, Russian Federation
Brian Hainey, Glasgow Caledonian University, UK
Sungkook Han, Wonkwang University, South Korea
Ville Isomottonen, University of Jyvaskyla, Finland
Mirjana Ivanovic, University of Novi Sad, Serbia
Jason J. Jung, Yeungnam University, South Korea
Samia Kamal, Oxford Brookes University, UK
Natalya Keberle, Zaporizhzhya National University, Ukraine
Vitaliy Kobets, Kherson State University, Ukraine
Oleksandr Kolgatin, H.S. Skovoroda Kharkiv National Pedagogical University, Ukraine
Christian Kop, Alpen-Adria-Universität Klagenfurt, Austria
Hennadiy Kravtsov, Kherson State University, Ukraine
Vladislav Kruglik, Kherson State University, Ukraine
Sergey Kryukov, Southern Federal University, Russian Federation
Vladimir Kukharenko, National Technical University "Kharkiv Polytechnic Institute", Ukraine
Nataliya Kushnir, Kherson State University, Ukraine
Vira Liubchenko, Odessa National Polytechnic University, Ukraine
Alexander Lyaletski, Taras Shevchenko National University of Kyiv, Ukraine
Dmitry Maevsky, Odessa National Polytechnic University, Ukraine
Frederic Mallet, Universite de Nice-Sophia Antipolis, France
Wolf-Ekkehard Matzke, MINRES Technologies GmbH, Germany
Heinrich Mayr, Alpen-Adria-Universität Klagenfurt, Austria
Mykola Nikitchenko, Taras Shevchenko National University of Kyiv, Ukraine
Tope Omitola, University of Southampton, UK
Olga Ormandjieva, Concordia University, Canada
Simon Parkinson, University of Huddersfield, UK
Vladimir Peschanenko, Kherson State University, Ukraine
Gary Pratt, Eastern Washington University, USA

Carlos Ruiz, playence, Spain
Abdel-Badeeh Salem, Ain Shams University, Cairo, Egypt
Wolfgang Schreiner, RISC, Johannes Kepler University Linz, Austria
Pavlo Serdyuk, Lviv Polytechnic National University, Ukraine
Vladimir A. Shekhovtsov, Alpen-Adria-Universität Klagenfurt, Austria
Mariya Shishkina, Institute of Information Technologies and Learning Tools of the National Academy of Pedagogical Sciences of Ukraine, Ukraine
Martin Strecker, IRIT, Paul Sabatier University, Toulouse, France
Ilias Tachmazidis, University of Huddersfield, UK
Olga Tatarintseva, Satelliz, Ukraine
Vagan Terziyan, University of Jyväskylä, Finland
Ville Tirronen, University of Jyvaskyla, Finland
Nikolay Tkachuk, National Technical University "Kharkiv Polytechnic Institute", Ukraine
Mauro Vallati, University of Huddersfield, UK
Leo Van Moergestel, Utrecht University of Applied Sciences, The Netherlands
Maxim Vinnik, Kherson State University, Ukraine
Paul Warren, Knowledge Media Institute, the Open University, UK
Vitaliy Yakovyna, Lviv Polytechnic National University, Ukraine
Yulia Nosenko (Zaporozhchenko), Institute of Information Technologies and Learning Tools of the National Academy of Pedagogical Sciences of Ukraine, Ukraine
Iryna Zaretska, V. N. Karazin Kharkiv National University, Ukraine
Grygoriy Zholtkevych, V. N. Karazin Kharkov National University, Ukraine

Additional Reviewers

Kalliopi Kravari, Aristotle University of Thessaloniki, Greece
Rustam Gamzaev, National Technical University "Kharkiv Polytechnic Institute", Ukraine
Eleftherios Spyromitros-Xioufis, Aristotle University of Thessaloniki, Greece
Emmanouil Rigas, Aristotle University of Thessaloniki, Greece

ITER 2015 Workshop

Workshop Chairs

Vitaliy Kobets, Kherson State University, Ukraine
Sergey Kryukov, Southern Federal University, Russian Federation
Sergey Mazol, Academy of Public Administration, Minsk, Belarus
Tatyana Payentko, National University of State Tax Service of Ukraine, Ukraine

Program Committee

Tom Coupe, Kyiv School of Economics, Ukraine
Dorota Jelonek, Częstochowa University of Technology, Poland
Ludmila Konstants, American University of Central Asia, Kyrgyz Republic
Sergey Kryukov, Southern Federal University, Russian Federation
Sergey Mazol, Academy of Public Administration, Minsk, Belarus
Marin Neykov, University of National and World Economy (UNWE), Bulgaria
Nina Solovyova, Kherson State University, Ukraine
Ekaterina Vostrikova, Astrakhan State University, Russian Federation
Alexander Weissbult, Kherson State University, Ukraine

MRDL 2015 Workshop

Workshop Chairs

Vladimir Kukharenko, National Technical University "Kharkiv Polytechnic Institute", Ukraine
Yulia Nosenko (Zaporozhchenko), Institute of Information Technologies and Learning Tools of the National Academy of Pedagogical Sciences of Ukraine, Ukraine
Hennadiy Kravtsov, Kherson State University, Ukraine

Program Committee

Olga Gnedkova, Kherson State University, Ukraine
Alexander Kolgatin, H.S. Skovoroda Kharkiv National Pedagogical University, Ukraine
Evgen Kozlovskiy, Kherson State University, Ukraine
Vladislav Kruglik, Kherson State University, Ukraine
Michael Sherman, Kherson State University, Ukraine
Maria Shishkina, Institute of Information Technologies and Learning Tools of the National Academy of Pedagogical Sciences of Ukraine, Ukraine
Tatyana Zaytseva, Kherson State Maritime Academy, Ukraine

SMSV 2015 Workshop

Workshop Chairs

Wolfgang Schreiner, RISC, Johannes Kepler University Linz, Austria
Mykola Nikitchenko, Taras Shevchenko National University of Kyiv, Ukraine
Michael Lvov, Kherson State University, Ukraine
Martin Strecker, IRIT, Paul Sabatier University, France

Program Committee

Anatoliy Doroshenko, Glushkov Institute of Cybernetics of the National Academy of Sciences of Ukraine, Ukraine
Louis Feraud, Paul Sabatier University, France
Alexander Letichevsky, Glushkov Institute of Cybernetics of the National Academy of Sciences of Ukraine, Ukraine
Alexander Lyaletski, Taras Shevchenko National University of Kyiv, Ukraine
Frederic Mallet, University of Nice Sophia Antipolis, France
Vladimir Peschanenko, Kherson State University, Ukraine

TheRMIT 2015 Workshop

Workshop Chairs

Vyacheslav Kharchenko, National Aerospace University “KhAI”, Ukraine
Elena Zaitseva, Žilina University, Slovakia
Bogdan Volochiy, Lviv Polytechnic National University, Ukraine

Program Committee

Mario Fusani, ISTI-CNR System and Software Evaluation Center, Italy
Vladimir Sklyar, National Aerospace University "KhAI", Ukraine
Iosif Androulidakis, Ioannina University Network Operations Center, Greece
Yuriy Kondratenko, Black Sea State University named after Petro Mohyla, Ukraine
Vitaly Levashenko, Žilina University, Slovakia
Dmitriy Maevskiy, Odessa National Polytechnic University, Ukraine
Vladimir Mokhor, Pukhov Institute for Modeling in Energy Engineering, NASU, Ukraine
Oleg Odarushchenko, Research and Production Company Radiy, Ukraine
Olexandr Gordieiev, University of Banking of National Bank of Ukraine, Kyiv, Ukraine
Yurij Ponochovny, Poltava National Technical University, Ukraine
Jüri Vain, Tallinn University of Technology, Estonia
Sergiy Vilkomir, East Carolina University, USA
Vladimir Zaslavskiy, Taras Shevchenko National University of Kyiv, Ukraine

Local Organizing Committee

Orest Lavriv, Lviv Polytechnic National University, Ukraine
Lyudmyla Novgorodska, Lviv Polytechnic National University, Ukraine
Leonid Ozirkovsky, Lviv Polytechnic National University, Ukraine
Oksana Soprunyuk, Lviv Polytechnic National University, Ukraine

Sponsors

Oleksandr Spivakovsky's Educational Foundation (**OSEF**, http://spivakovsky.fund/) aims to support gifted young people, outstanding educators, and also those who wish to start up their own business. OSEF activity is focused on the support and further development of educational, scientific, cultural, social and intellectual spheres in the Kherson Region of Ukraine.

DataArt (http://dataart.com/) develops industry-defining applications, helping clients optimize time-to-market and minimize software development risks in mission-critical systems. Domain knowledge, offshore cost advantages, and efficiency – that's what makes DataArt a partner of choice for their global clients.

Lviv Polytechnic National University (http://www.lp.edu.ua/en) is the largest technological university in Lviv. Since its foundation in 1844, it was one of the most important centres of science and technological development in Central Europe. Presently, the university comprises 16 institutes where students from Ukraine and other countries are enrolled in 64 bachelor, 123 master, and 99 PhD programmes.

Logicify (http://logicify.com/) is an outsourcing company providing software development services. Compay helps customers with issues and projects involving software. Logicify has been working in a variety of industries and fields, including telecom, video sharing, social media, insurance. It has several teams with specialized skills in different technologies that can relate to specific industries.

Organizers

Ministry of Education and Science of Ukraine
http://www.mon.gov.ua/

Lviv Polytechnic National University, Ukraine
http://www.lp.edu.ua/en

University of Huddersfield, UK
http://www.hud.ac.uk/

Alpen-Adria-Universität Klagenfurt, Austria
http://www.uni-klu.ac.at/

Kherson State University, Ukraine
http://www.kspu.edu/

Taras Shevchenko National University of Kyiv, Ukraine
http://www.univ.kiev.ua/en/

V.N. Karazin Kharkiv National University, Ukraine
http://www.univer.kharkov.ua/en

Zaporizhzhya National University, Ukraine
http://www.znu.edu.ua/en/

Institute of Information Technologies and Learning Tools
of the National Academy of Pedagogical Sciences
of Ukraine, Ukraine; http://iitlt.gov.ua/en/

DataArt Solutions Inc., Russian Federation
http://dataart.com/

Table of Contents

TheRMIT Workshop Papers

Rigorous Semantics and Refinement for Business Processes (Abstract)*

Klaus-Dieter Schewe[1,2]

[1] Software Competence Center Hagenberg, Austria, kd.schewe@scch.at
[2] Johannes-Kepler-University Linz, Austria, kd.schewe@cdcc.faw.jku.at

Keywords. Business process model, Abstract State Machine, semantics, refinement, exception handling

ICTERI Key Terms. Mathematical Model, Methodology, Formal Method, Process, Integration

For the modelling of business processes it is necessary to integrate models for control flow, messaging, event handling, interaction, data management, and exception handling. In principle, all common business process models such as BPMN [14], YAWL [13], ARIS [11] or S-BPM [6] follow such an approach. Though it is claimed that the models have already reached a high level of maturity, they still lack rigorous semantics as pointed out in [1, 5, 15]. Furthermore, quite a few aspects such as data management, interaction and exception handling have only been dealt with superficially as pointed out in [12].

The first concern regarding rigorous semantics has been discussed in detail by Börger in [2] for BPMN, which led to an intensive investigation of BPMN semantics on the grounds of Abstract State Machines (ASMs, [4]), in particular for OR-synchronisation [3]. The monograph by Kossak et al. defines a rigorous semantics for a large subset of BPMN leaving out some ill-defined concepts [8].

The second concern can be addressed by means of horizontal refinement. On grounds of ASMs necessary subtle distinctions and extensions to the control flow model such as counters, priorities, freezing, etc. can be easily integrated in a smooth way [12]. Conservative extensions covering messaging can be adopted from S-BPM [6], while events in BPMN have been handled in [7]. For the event model it is necessary and sufficient to specify what kind of events are to be observed, which can be captured on the grounds of monitored locations in ASMs, and which event conditions are to be integrated into the model. Extensions concerning actor modelling, i.e. the specification of responsibilities for the execution of activities (roles), as well as rules governing rights and obligations lead to the integration of deontic constraints [10], some of which can be exploited to simplify the control flow [9]. In this way subtle distinctions regarding decision-making responsibilities in BPM can be captured.

* The research reported in this paper was supported by the Austrian Forschungsförderungsgesellschaft (FFG) for the Bridge Early Stage project "Advanced Adaptivity and Exception Handling in Formal Business Process Models" (adaBPM) under contract **842437**.

In the talk a glimpse of the rigorous, ASM-based semantics for business processes is presented. The focus is on the control flow with specific emphasis on priority handling. This is followed by a discussion of horizontal refinement focusing on the introduction of disruptive events and associated exception handling. A simplified example capturing the effects of external change to a running process is used for illustration.

References

1. Abramowicz, W., Filipowska, A., Kaczmarek, M., Kaczmarek, T.: Semantically enhanced business process modelling notation. In: Hepp, M., et al. (eds.) S-BPM. CEUR Workshop Proceedings, vol. 251. CEUR-WS.org (2007)
2. Börger, E.: Approaches to modeling business processes: a critical analysis of BPMN, workflow patterns and YAWL. Software & Systems Modeling 11(3), 305–318 (2012)
3. Börger, E., Sörensen, O., Thalheim, B.: On defining the behavior of OR-joins in business process models. Journal of Universal Computer Science 15(1), 3–32 (2009)
4. Börger, E., Stärk, R.: Abstract State Machines. Springer-Verlag, Berlin Heidelberg New York (2003)
5. Dumas, M., La Rosa, M., Mendling, J., Reijers, H.A.: Fundamentals of Business Process Management. Springer (2013)
6. Fleischmann, A., et al.: Subject-Oriented Business Process Management. Springer-Verlag, Berlin Heidelberg New York (2012)
7. Kossak, F., Illibauer, C., Geist, V.: Event-based gateways: Open questions and inconsistencies. In: Mendling, J., Weidlich, M. (eds.) Business Process Model and Notation, Lecture Notes in Business Information Processing, vol. 125, pp. 53–67. Springer, Berlin, Heidelberg (2012)
8. Kossak, F., et al.: A Rigorous Semantics for BPMN 2.0 Process Diagrams. Springer-Verlag (2014)
9. Natschläger, C., Kossak, F., Schewe, K.D.: BPMN to Deontic BPMN: A trusted model transformation. Journal of Software and Systems Modelling (2015), to appear
10. Natschläger-Carpella, C.: Extending BPMN with Deontic Logic. Logos Verlag, Berlin (2012)
11. Scheer, A.W.: ARIS - Business Process Modeling. Springer, Berlin, Heidelberg (2000)
12. Schewe, K.D., et al.: Horizontal business process model integration. Transacions on Large-Scale Data- and Knowledge-Centered Systems 18, 30–52 (2015)
13. ter Hofstede, A.M., et al. (eds.): Modern Business Process Automation: YAWL and its Support Environment. Springer, Heidelberg (2010)
14. Weske, M.: Business Process Management. Concepts, Languages, Architectures. Springer (2012)
15. Wong, P.Y., Gibbons, J.: A process semantics for BPMN. In: Liu, S., Maibaum, T., Araki, K. (eds.) Formal Methods and Software Engineering. Lecture Notes in Computer Science, vol. 5256, pp. 355–374. Springer, Berlin Heidelberg (2008)

Smart Learning Environments: a Shift of Paradigm

David Esteban[1]

[1] TECHFORCE, Vía Augusta, 2bis planta 5ª E-08006 Barcelona, Spain

desteban@techforce.eu

Abstract. The incorporation of Information and Communication Technologies (ICT) as a supporting mechanism in educational processes has already been proved as an important driver in reinforcing both teaching and learning. The extensive development of Learning Management Systems (LMS), software platforms aimed at supporting and articulating e-learning, education courses and training programs, is already backed by a relevant ICT industry, with significant market penetration. The emergence of the new concept of Smart Learning Environments (SLEs) is shifting the main focus of LMSs on courseware towards a more efficient and effective approach focused on teaching and learning processes, thus in the students themselves and in the teachers as key players. The evolving concept of SLEs encompasses blending educational technologies with appropriate considerations and guidance developed by pedagogical and educational neuroscience domains, thus opening up room for interesting scientific and technological challenges.

Keywords. Information and Communication Technology, Learning Management System, Smart Learning Environment

Key Terms. InformationCommunicationTechnology, TeachingMethodology, TeachingProcess, Environment

Systematic Business Process Modeling in a Nutshell

Heinrich C. Mayr[1]

[1]Alpen-Adria-Universität Klagenfurt,
Universitätsstrasse 65-67 Klagenfurt, 9020, Austria
Heinrich.Mayr@aau.at

Abstract. In-depth business process management is crucial for any institution and enterprise in a competitive world. Although this insight is by no means new, the daily practice draws another picture: Certainly, many enterprises have defined their overall strategy including IT issues at least roughly, and, based here on, have documented their business processes somehow. Rarely however, do they manage their business processes comprehensively in the sense of covering analysis, design, measurement, continuous optimization, and IT support.
The key prerequisite for allowing such comprehensive handling of business processes is to describe these processes transparently and completely, using a modeling language that is appropriate for the particular context including all stakeholders concerned.
The aims of this tutorial, therefore, are threefold: (1) the participants will learn about the fundamentals of business processes and their contexts; (2) the key features of popular business process modeling languages like Adonis and BPMN; and (3) guidelines for selecting an appropriate modeling approach including the customization to the given environment.
Intended audience: Practitioners and researchers who are interested in a systematic approach to business process management, and have basic knowledge in modeling and information systems engineering.

Keywords. Business process fundamental, business process context, business process modeling language, selection of the modeling approach, customization

Key Terms. Process, ProcessPattern, Technology, Methodology, Model

Using ICT in Training Scientific Personnel in Ukraine: Status and Perspectives

Aleksandr Spivakovsky, Maksim Vinnik and Yulia Tarasich
Kherson State University, 27, 40 rokiv Zhovtnya St., 73000 Kherson, Ukraine
{Spivakovsky, Vinnik, YuTarasich}@kspu.edu

Abstract. Today an enormous amount of problems in building a system of efficient education and science is on the discussion agenda in Ukraine. A decrease in the number of scientists in the country has been observed in the last 15 years. At the same time, the amount of postgraduate students and people aiming at obtaining their doctorate is increasing. Notably, similar indicators are also observed in the majority of post-soviet countries. One complicating factor is that the system of scientific personnel training in Ukraine is very restrictive and closed. The proportion of research results published using a free access scheme to the overall bulk of publications is still very small, in particular if compared to the level of ICT development. Therefore, a major part of the publications still remains inaccessible from the outside. In this study we investigate the openness and accessibility of the preparation of the academic staff in Ukraine. As a result we come up with a proposal of requirements to the ICT infrastructure in this area.

Keywords: Information and communication technology, Education and learning process, ICT infrastructure, Open Science.

Key Terms: ICT Infrastructure, Research.

"If it's not on the Web, it doesn't exist at all"
Sarah Stevens-Rayburn & Ellen N. Bouton, 1997

1 Introduction

The main catalyst for socio-economic development of a state potential is the ability to create, collect, and effectively manage knowledge that is comes out from the best scholarly research practices. The countries which have made it to their development strategy and implemented the effective interaction with the business enjoy TOP ratings in the World rankings. In the age of information technologies, it takes one not years, but rather days to bear the bell of scientific research and excel the competitors. The companies which are the first in the market are more likely to benefit from a positive effect caused by the introduction of new knowledge. Globalization is adjusting the cooperation between science and industry. More and more funds are invested in scientific research and development to capture the leadership in the market. A modern country's development is stimulated by the transition from a resource-based economy to hi-tech. There is an opportunity to create “intellectual

dollars" without any resource, but people. The results of intellectual work become a hard currency. For example, Japan, though it had no natural resources, managed to become the leader in world's economy. The monetary value of the biggest hi-tech (IT) companies is at a scale of the budgets of some developed countries (Apple – $ 711 billion, Microsoft – $ 349 billion, Google – $ 365 billion).

The Open Science (OS) movement gains popularity in the world of clerisy, aiming to make research results and source data accessible to public at all levels. However, there is a conflict between the desire of scientists to have access to shared resources and make profit by using these resources [1]. In recent years, many governments try to impose the policy of openness regarding scientific knowledge, especially, if it is funded with public money. One way is the enforcement of providing open access to the results of all research projects performed at public expense. An indicative example is the US, which grant annually about $ 60 billion for research. In 2008, the US Congress imposed the obligation to grant free access in a year after the first publication to all the research papers based on the studies conducted by the National Institute for Health (which receives circa the half of the total public funding for science). Similar measures are now considered by many other countries.

Today, a lot of research in Ukraine is devoted to the problems of higher education and, in particular, the use of ICT for training students, creating information and communication environments in the universities, etc. However, in the scholarly literature insufficient attention is paid to the development of information and communication models of interaction with ICT in academic staff training. Moreover, today we are talking about the need for openness and accessibility of scientific activity, whereas a substantial part of the scholarly output never reaches its reader within and even more outside the professional academic community. This problem is particularly acute in the post-soviet countries. Regionalism of entire areas in science, convention, low connection with contemporary scientific trends, low level of foreign language knowledge by scientists, lack of self-developing scientific community, low competition with other countries, lack of motivation, poor funding, brain drain, and a number of other factors result in the continuing archaism of scientific brainpower training in Ukraine.

Scientometrics is rapidly developing nowadays. Using information technology allows creating new services for the development of scientific and research activity. Many global companies invest billions of dollars in services to support research activity, thereby creating a serious market not for the research results but for the research process support. Herewith the trend shifts toward commercial projects. The examples of such companies are Apple, Microsoft, Google, Elsevier, Thomson Reuters, not to mention many others. The most outstanding services with rapidly growing impact are Google Scholar, Scopus, Orcid, Academia.edu, Research Gate, Mendeley, arXiv.org, cs2n, Epernicus, Myexperiment, Network.nature, Science-community. These services contribute to satisfying the needs of the scientific community. In fact, these positively influence scientific and technical progress and create a new paradigm of scientific research. A big number of the recently created scientometric services allow assessing the relevance of the research results by a scientist, the number of his publications, citations, storage, etc. Having these measurements at hand opens up new opportunities and prospects. Our time is characterized by the high rates of the accumulation of new knowledge, in particular in

the form of research results. Provided that the integration of research activities is currently (and naturally) low, a huge amount of scientific and research information falls out of search visibility and accessibility. Information technology is the only way to arrange and create effective search tools for acquiring the necessary knowledge. The objective of our research is to investigate the transparency of specialized scientific bodies and offer the vision of their supporting ICT infrastructure. Accordingly, the rest of the paper is structured as follows.

Present article includes such sections, as description of the methodological and experimental parts (2-4), discussion of basic components of DC's ICT infrastructure and main ways and methods of their realization (5).

2 Related Work

David [2] mentions that the goal of Open Science is to do scientific research in a way that facts and their distribution is made available at all the levels of the concerned public. The same article states that the movement arose in the XVII century. Due to the fact that the public demand for access to scientific knowledge has become so large that there was a need for a group of scientists to share their resources with each other, so that they could conduct research collectively [2].

The term E-Science (or eScience) was proposed by John Taylor, the Director-General of the United Kingdom Office of Science and Technology in 1999 and was used to describe a large funding initiative, starting from November 2000. E-Science has been interpreted more broadly since as "the application of computer technology to the implementation of modern scientific research, including training, experimentation, accumulation, dissemination of results and long-term storage and access to all materials obtained through the scientific process. These may include modeling and analysis of facts, electronic/digitized laboratory notebooks, raw materials and built-in data sets, handwritten production and design options, preprints and print and/or electronic publications"[3].

Koichiro Matsuura, the President of UNESCO, wrote in his preface to [4]: "Societies that are based on the knowledge will need to share them to keep their human nature".

In 2014, the IEEE eScience community proposed a condensed definition [5]: "eScience encourages innovation in collaborative, computationally or facts intensive research in all the disciplines throughout the research life cycle".

Michael Nielsen, a physicist and propagandist of Open Science, colorfully describes in [6] the way the new instruments need to look like to facilitate the dissemination of the culture of cooperation and openness among scientists. One of such tools exists now. This is arXiv – a site that allows physicists to publish preprints of their works before the official publication of the article. This promotes to get in faster feedback and to disseminate the new discoveries. Nielsen also acts for publishing not only conclusions, but all the original data – this is the thing physicists have been dreaming of for a long time. Journals could help them do that if they wanted to [6].

The peer review system for scientific papers on one hand offers an opportunity to obtain a (preliminary) critical assessment of a manuscript, but on the other hand it slows down the publication of research results. In this system, a review process is rarely accomplished in less than a month. The reviewers often request authors to revise some parts of the material or conduct additional studies. As a result, the time before the publication stretches for about six months or more. However, Michael Eisen, the co-founder of the Public Library of Science (PLoS), mentioned that according to his experience the "most serious incompletes are detected only after the article is published." The same applies to other scientific works, including dissertations for a degree [7]. The cases are known in history when after many years after the defense a person was divested a degree and even was fired after the examination of his work regarding qualitative or even plagiarism.

Tugo Pagano and Maria Alessandra Rossi suggest [8] that politics aimed at overcoming the disadvantages of excessive privatization of knowledge can play an important role in stimulating the economy. Efforts should be focused to maintain and enhance the role of open science. The institutions of open science have allowed the flourishing of industrial development from the beginning, and should have a much more important role in the architecture of the future post-crisis global economy. This can be achieved through the institute of World Research Organization (WRO) which can master some of the benefits of open science to overcome the well-known free rider problem associated with contributions to the last.

In 2004, the research group Laboratorio de Internet from Spain, which studies educational and scientific activities on the Internet, started the Webometrics (www.webometrics.info) project with the aim to rate University web sites. The subject of their analysis is the university domain. Webometrics researchers emphasize that the presence of a university website allows to simplify the publication of scientific works by faculty and research staff, compared to the publication in print, and also provides the information the fields of their professional activities. Online publications are much cheaper than paper publications and have broader potential audience. Publishing online facilitates to broadening the access to academic resources for scientific, commercial, political, and cultural organizations both from within a country and abroad. The rating scale is based on the four criteria that take into account the entire Web data within the university domain: Visibility, Presence, Openness, and Excellence. Each criterion has a weight corresponding to its importance [9].

The report by UNESCO on information technology in education [4] shows that in Ukraine there is a “rapid advancement of ICT into the sphere of education, which needs continuous improvement in the efficiency of use of the new ICT in the educational process, timely updates of educational content, and an increase in the quality of ICT training”. However, there are some problems which are primarily associated with the low psychological, methodological, and pedagogical readiness of teachers to the rapid changes in information technology.

The issue of the openness of an education system and science often comes up in relation to international research funding instruments, such as Tempus, Erasmus, and others, and related projects. Every year, they attract the attention of many Ukrainian and foreign universities, research organizations and structures.

In 2006-2008 our Kherson State University (KSU) participated in the following European projects: Tempus TACIS CP No 20069-1998 “Information Infrastructure of Higher Education Institutions”; Tempus TACIS MP JEP 23010-2002 “UniT-Net: Information Technologies in the University Management Network”; US Department of State Freedom Grant S-ECAAS-03-GR-214(DD) “Nothern New York and Southern Ukraine: New Partnership of University for Business and Economics Development”, which resulted in the development and implementation of scientific and management processes of analytical information systems and services. More detailed information can be found in the articles by G. Gardner [10], V. Ermolayev [11], A. Spivakovsky [12].

The results on the interrelation of ICT and educational process and the influence of ICT on professional and information competencies of the future university graduates have been presented in our previous publications [13, 14]. The authors have also conducted the investigation of the technical component of the feedback services implementation in KSU [15] and their impact on the preparedness of the students to use ICT for educational and non-educational purposes, and forming the ICT infrastructure in a higher educational institution [16, 17].

3 Experimental Settings

Today, Ukraine possesses a historically established system of scientific training. The foundations of this system were laid in the Soviet Union. This system is very similar to the system of post-soviet countries.

According to the State Statistics Service, 2011 [18], Ukraine had 14 895 “doctors of science” and 84 979 “candidates of science” (the analog of a PhD) covering arts, legal studies, and sciences. Among them 4 417 doctors and 16 176 candidates of science work in sciences. In addition, as reported by the ”Voice of Ukraine” newspaper, the National Academy of Sciences of Ukraine employs today 2 564 doctors and 7 956 candidates of science [18].

In the last 19 years the number of researchers in Ukraine, decreased by more than 100 thousand people, while the number of graduate students increased by almost 2 times (Fig. 1 shows an example). The trend similar to the decrease in the research staff members can be observed in the numbers of domestic research and development organizations (Fig. 2 shows an example).

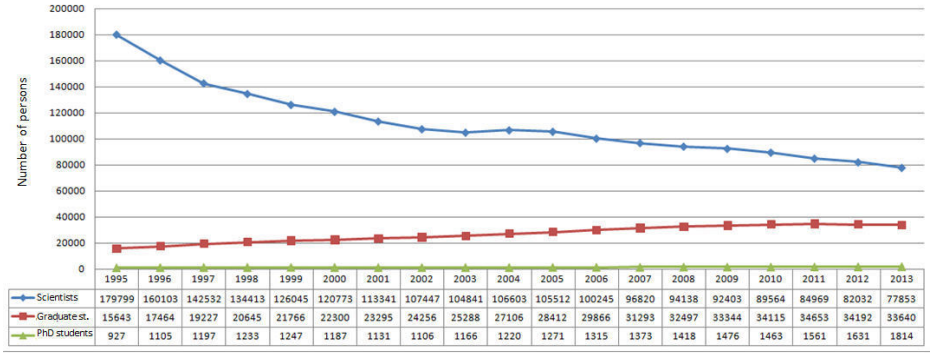


Fig. 1. The dynamics in the numbers of research staff, PhD students, and university graduates in Ukraine (1995-2013).

In Ukraine there are 988 Dissertation Committees (DC) [19]. DC are the expert councils in different scientific domains which form the National organizational infrastructure, accepting candidate and doctoral dissertations for examination, doing the expertise, hosting the defenses of dissertations, and further awarding advanced academic degrees. The aim of this infrastructure is to foster the development of the innovative elite of Ukraine which is considered as a driving force for scientific and technological progress.

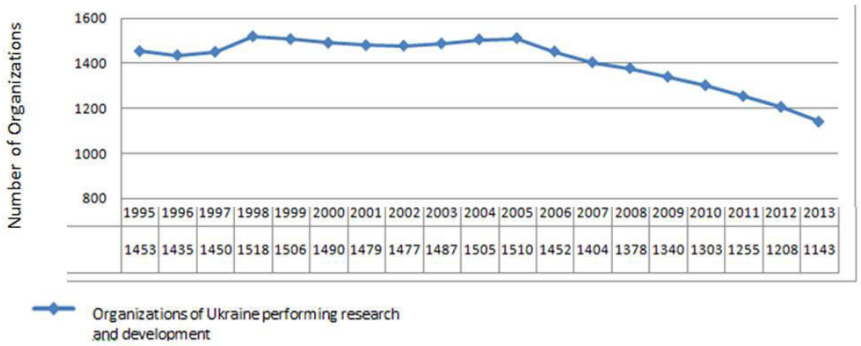


Fig. 2. The dynamics in the numbers of Ukrainian organizations performing research and development (1995-2013).

Given the importance of the DC infrastructure, the foci of this study are to:

- Assess the openness and accessibility of the preparation of academic staff in Ukraine within the system using the DC.
- Specify the requirements for the construction of the ICT infrastructure in this area We will analyze the performance of DC based on the following principles:

1. The availability of information;
2. Openness;

3. Weight;
4. Scientific;
5. Social significance.

The research into the current state of the system of interaction with ICT of the DC, the Higher Attestation Commission of Ukraine, and graduate students is impossible without the analysis, comparison and synthesis, abstract approach to the definition of the basic patterns of the use of information technologies, and logical approach to the description of possible implementations of innovative teaching methods. Hence, the study of this issue requires the use a carefully designed combination of exploratory, empirical, and statistical methods. Therefore, several methods are used:

- Exploratory – the analysis, synthesis, comparison, generalization and systematization of relevant information acquired from psychological and educational literature legal documents, standards and information resources. These sources are consulted and further generalized to define the essence of the information competency of university students and assess the theoretical and methodological bases of information competency formation. Pedagogical modeling is employed to build the model of informatics competency.
- Empirical – questionnaires, surveys, testing, and self-esteem; pedagogical experiments are used to test the hypotheses of the study
- Statistical – the methods of mathematical statistics are employed to determine the reliability of the results on the basis of quantitative and qualitative analysis of the empirical data

The analysis of the public (available on the Internet) information on the availability of data on DC, and collecting the opinions of graduate students using a questionnaire on the use of information technology in their dissertation projects are the main research methods.

Considering that the DCs function as university bodies, such sites as Top 100 universities in the World, Top 10 European universities, Top 50 universities in Russia, Top 25 universities in Poland, Top 10 universities in the USA, Top 15 universities in UK, Top 20 universities in Asia [20], Specialized DC of Ukraine were the object of information analysis. Overall, 300 university sites were analyzed in the reported research.

The study of the current status the use of ICT to support the activities of DC the following assessment aspects:

1. The availability of a web site for a DC and its analysis;
2. The degree of openness of the information provided for a DC: information about the members, dissertation abstracts, theses, etc;
3. Information security;
4. The existence of DC pages in social networks;
5. The availability of a feedback service.

Let us consider in more detail each of the assessment aspects.

1. While exploring the web sites of universities regarding the availability of information about the activities of the respective DC, we have selected to use the following four criteria:

- A university web site provides the information on the DC and a link to its own website
- A DC does not have a separate web site, but it has a page on the university web site
- A University website provides a brief information about the DC
- There is no information about the DC neither on the university website nor in social networks

2. The openness to the information about a DC for public:

- Any Internet user can see the information
- A user can view the data only after registration on the web site
- Only the staff and students of the university can see the information

3. Feedback facilities:

- Providing a contact phone number;
- Providing a contact e-mail address(es);
- Providing the list of contact persons;
- Providing the Skype ID for contacts;
- Providing the schedule of DC works.

4. The availability of information (pages) in social networks:

- Due to a substantial impact of social networks on the communication among people today, it has been decided to account for the relevant indicators in our study an analysis of the availability of information about DC: the availability of accounts or groups in social networks such as Vkontakte, Facebook, Google+, Twitter
- To analyze the availability of video records of defense meetings the analysis of the YouTube content relevant to a DC has been also undertaken

5. The technical characteristics of DC web sites used in our study are detailed in Table 1.

Table 1. Technical criteria for the analysis and evaluation of DC websites.

Criterion	Description
Number of Web Pages	The number of DC-relevant pages on a web site is the indicator influencing the ranking of the site in search results.
Frequency of Updates	The frequency of updating information about the activities of a DC is analyzed using the scale: weekly, monthly or annually
Authentication System	The main elements and authentication mechanism are analyzed under this aspect.
Usability	The assessment of ease of use and operation of the system is done under this aspect, namely how well, clearly and correctly the interface is implemented and web site is structured. It is also assessed if a user can quickly find the information he or she needs. We conducted a brief analysis of layout. We also checked the availability dynamic elements and search functionality.

Platform	The web sites were categorized as implemented using CMS and hand-coded.
SEO	Under this aspect the ranking of a web site by search engines for specific user requests was analyzed.
Validity	Under this aspect we looked at the number of errors found by the web site validator (http://validator.w3.org/).
Multimedia content	A study on the website of the libraries of audio and video recordings protections scientific papers, photographs, etc.

The questionnaire which has been used to survey the use of ICT by graduate students in their preparation to defense consisted of 3 components:

- Quantitative indicators of the use of ICT by graduate students in the process of working with their DC.
- The availability of training courses for the use of ICT in the preparation to defense
- The readiness of the subjects to authorize the open storage of their research results (articles, theses, dissertations) and review materials such as audio, video, etc.

4 Experimental Results

The result of the analysis of the websites of the universities of Ukraine regarding the information on DC, personal web pages and sites of DC members is pictured in Fig. 3. Only 9% of the reviewed DC have their own web sites. 84% of DC related information can be found on University web sites, taking into account that full information concerning the DC activities has been found only for 47% of the reviewed DC. 7% of the reviewed DC have no presence on the Internet. These results pinpoint the major problems in the transformation of the contemporary Ukrainian scientific community into the Open Science community.

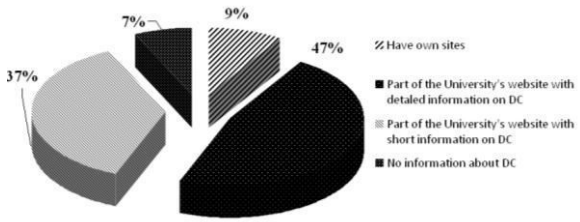


Fig. 3. The availability of DC related information on the web sites.

Only 4 DC web sites exploit a user authentication functionality distinguishing user roles. So, it can be stated that only 1% of the reviewed DC have created some ICT based prototypes for the interaction between the applicants and the Ministry of Education.

About 30% of the reviewed DC update the information on their web sites every week, whereas 51% of the information on these sites is updated several times per year

(Fig. 4 shows an example). Consequently, the question arises on the reliability and relevance of this information.

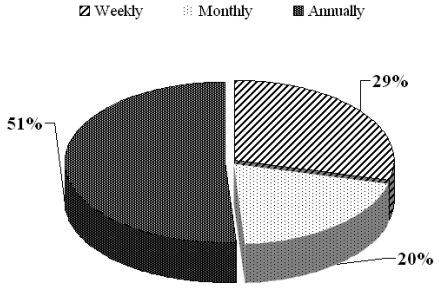


Fig.4. Frequency of updates.

As per the information on the reviewed web sites, the DC have no means to track scientometric indicators of the members of the DC, the candidates for a degree, and persons that had defended their theses in a particular DC, not to mention the presence of analysts defended dissertations and access to them, which makes the qualitative assessment of their activities impossible. 32 websites have usability problems in terms of the ease of use of their interfaces and poorly implemented site (keyword-based) search functionality. The latter is implemented on only 27 of the reviewed resources. Only 17 of the examined web sites provide the information on or references to resources like a “library”.

Regarding the minimally present contact information of a DC (a phone number, address, contact person name, document templates), it is provided only on 4 of the reviewed web sites. Moreover, the contact phone number is mentioned only on 2 of them. Thus, in order to find the information a DC of relevance to a PhD project, one should get their list and addresses in the Ministry of Education and Science of Ukraine (where one also needs to go) and search for a relevant DC at the specified address. This is only the first problem in the application process. The required documents have also to be submitted to a DC by coming in person, since there is not a single web site that allows you to exchange the information and documents with a DC in the process of registration, filing and review of the thesis and so on.

The results of the review of the availability of information about Ukrainian DC in social networks are shown in Fig. 5.

As can be seen in Fig. 5, 14 DC have a personal group or page in Vkontakte, 11 – in Facebook, 7 – in Twitter and 4 – in Google+. It is also important that YouTube is used, though to a small degree. So, a certain degree of openness of our science may be noted, in particular the openness of the preparation of the scientific staff.

The analysis of quantitative indicators of the use of ICT by graduate students for working with a DC is shown in Table 2.

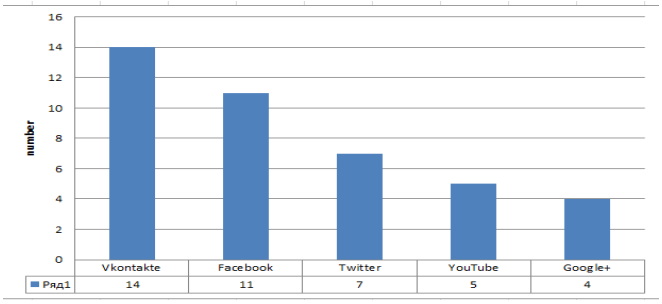


Fig. 5. The use of social networks in the work of DC.

The study reveals that only 2-3% of the respondents know what is a scientific database (SDB) or a citation index, 7% use these systems from time to time to find the necessary information, and only 4% have their own profiles in such scientometric systems and databases like Scopus, Google Scholar, Mendeley, RSCI or others. It is important that the majority of the respondents are not interested in creating their own profiles in such scientometric systems. The main reason for that is the lack of recognition of their utility. Moreover, some of the profiles were created directly by the organizations where scientists work, or automatically by the systems that store their scientific articles. Thus the majority of respondents did not know whether they have a profile in any of the systems, whether these exist or not.

80% of respondents do not think much about how their scientific publications are stored –in a paper or electronic form, and they believe that it is not of great importance. Thus, the majority of publications are going out of press in a paper form and are not further digitized – so remain unavailable to the scientific world.

Table 2. Quantitative indicators of the use of ICT by graduate students for working with a DC.

		Do not Use	Rarely Use	Always use
Use of the Internet to search for information about DC	Working with DC website	80%	15%	5%
	Search of information about the members of DC in SDB	93%	5%	2%
	Own profiles in SDB	95%	4%	1%
	Work with electronic repositories (theses and abstracts)	40%	50%	10%
Use of email		30%	40%	30%
Use of Skype		84%	10%	6%

93% of respondents answered negatively about attending any course (or lectures) to get prepared for the use of ICT in their dissertation project (SDB, repositories, etc.).

Analyzing the readiness to the open storage of research results (articles, theses, dissertations) and materials of dissertation defense such as audio or video, we observes the following:

1. The majority of the respondents (80%) support the publication of electronic copies of their scientific papers on the Internet, but at the same time consider it unnecessary and inconvenient. Further, all the respondents point out that the Ministry of Education and Science of Ukraine (MESU) has the publication requirements (regarding the number of papers and form of publication, paper or electronic) to qualify for a degree which do not motivate providing open access. MESU requires that a qualified candidate has 5 publications at the MESU approved venues, one of which can only be published in an electronic edition and another one in an international or indexed international SDB. Thus, none of the applicants target to publish the electronic copies of their papers on the Internet. In some cases, this problem is solved by posting electronic copies on a digest web site or putting these into an electronic repository of a scientific institution of the applicant. Otherwise the articles remain inaccessible to the outside world.
2. The Problem with open access to the protected dissertations and abstracts is identical to the previous. In addition, the human factor needs to be taken in consideration. Providing free access to abstracts or theses means making these open for further examination after publication, hence the increase of the author's responsibility for its contents and quality. Therefore, open storage of scientific work of this type stimulates quality improvement. We see it in the results of the evaluation of the respondents ' answers to this question. Notably, 80% of the respondents agree that the understanding that their work could be read by any other scientist clearly affects the quality of publications.

As an example, let us compare the quantities of the full versions of theses and abstracts stored in the repositories in Ukraine to numbers in the repositories in Germany, Great Britain, and Spain (top 30 repositories of each country rated by Webometrics, http://www.webometrics.info, were examined) – see Table 3. Ukraine has 38 repositories in total while having more than 400 universities.

Table 3. Numbers of dissertations and abstracts in open access repositories.

	UKRAINE	**GERMANY**	**UK**	**SPAIN**
Dissertation	1858	71656	16724	3586
Abstract	3532	22882	23617	18582

Only 15% of the respondents agree that online video protection is useful, 30% – to deposit their audio and video files providing open access, while the remaining 45% believe that audio and video recording is unnecessary or even harmful as it bothers and disturbs focusing on the defense talk. To the question "if they would like and are ready to use specialized systems to work with a DC and MESU" 90% of the respondents gave a positive answer. The most significant motive to this answer is potential reduction of time and financial expenses for data processing (sending and receiving documents, access to the proper information and so on).

5 Our Vision of an ICT Infrastructure for a DC

As experimentally proven above, the effective implementation of the elements of OS must assume the existence of an appropriate ICT infrastructure as a scientific and educational system as a whole and its component parts (schools, universities, DC, and others) in particular.

The main elements of the ICT infrastructure of OS are researchers (academic staff), data and processes.

Speaking of ICT infrastructure DC we can determine its components as follows:

- Researcher – the applicants, the members of a DC, the employees of MESU, and other users of the system have access to relevant information and participate in information processing, communication, and computing processes
- Data – information about the work of DC, their employees, applicants, archives of theses, scientific publications, etc. as a tool to open exchange, recombination, and reuse are the important components of the infrastructure;
- Process – the procedures, services, tools, and methodologies that are used to collect, perform the transformation, analysis, visualization and storage of data, build models and simulations. The management of these processes is done both on the side of users (researchers) and of the specialized services and systems.

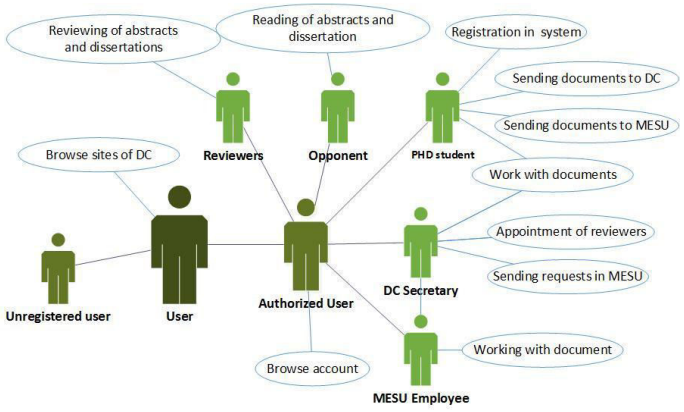


Fig.7. User roles and their main features.

As pictured in Fig. 7, all the user roles have both generic and specific abilities in using the system. All roles can retrieve publicly available information while working with documentation is allowed only to certain roles.

The workflow of the system is presented in Fig. 8 and proposes almost complete automation of all communication processes. It should be noted that the implementation of a similar service involves not only the functionality described above, but also the implementation of some add-ons and extra features. One of the additional features of interest is related to solving the problem of retrieving the information about the available DC discussed above. The task of collecting correct and complete scientometric data regarding the DC members, candidates, and

graduates is of particular importance and interest. It is difficult to compile by hand a report for an individual DC based on the scientometric information even if all the mentioned actors have their profiles, say at Google Scholar. The task of reporting about all relevant DC, or the graduates interested in applying to a relevant DC, is even more complicated.

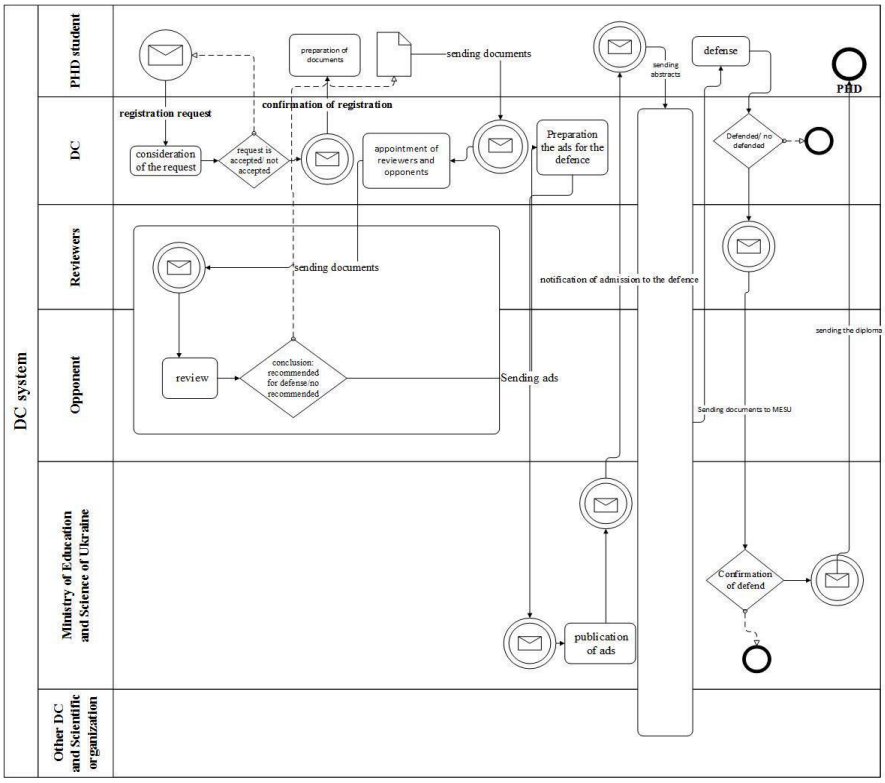


Fig. 8. Algorithm of DC system work.

We currently possess a number of modules relevant for the solution of this problem. For example, the problem outlined above may be solved using our publication.kspu.edu service. The main task of this service is automating the collection and processing of information on scientometric indicators of scientist retrieved from his or her Scopus and Google Scholar profiles and building consolidated ratings for departments, institutes, and universities.

The service provides the possibility to generate the required scientometric indicators and ratings on the DC web site, hence offering an additional degree of openness of their activities. In addition, we assume that this form of presenting information contributes to establishing vigorous competition in the research staff training market, and therefore has an impact on the quality indicators.

An electronic repository of a DC is also a mandatory component of the system which store all theses and abstracts in electronic form and, if possible, in the public domain.

The competition on the market puts in front the requirement of using social networks in the activities of a DC. The results of the study presented in Section 4 provide clear evidence about their rare use. We believe that the inclusion of the use of social networks in the workflow will provide a valuable addition to the information and communication services of the architected infrastructure.

6 Concluding Remarks and Future Work

Building a system of efficient education and science in Ukraine today is complicated by many serious problems. In the last 15 years we observed a decrease in the number of scientists in the country. At the same time, the numbers of postgraduates and doctorates are increasing.

A system of training of scientific personnel in Ukraine is among the most restrictive and closed ones in the world. A similar trend is observed in the majority of post-soviet countries. The proportion of scientific research results published under a open access is still very small compared to the level of ICT development. The main part of research results still remains inaccessible for an external users.

The use of ICT in training scientific personnel and representing the results of their research appears to be extremely weak. The preparation, protection and storage of information is done without using ICT, therefore it requires significant time and resource.

To partly overcome some of the problems, we propose a concept of the ICT infrastructure for the interaction of researchers, a DC, and the Ministry of Education and Science of Ukraine. The main elements of this infrastructure are the following: the web sites and services for supporting the applicants to a DC; the services and systems of interaction between a DC and the Ministry of Education and Science of Ukraine; electronic data storages for publications, theses, and abstracts; decreasing of time spent on a research process without harming its quality; additional expertise; transparency and credibility of research; building qualitatively new communications between scientists; ability to obtain research information swiftly and in required forms (especially government); fighting corruption (decreasing human factor in DC activities).

An important and influential part of establishing an effective process for training scientists is training researchers to use this process and respective tools based on ICT. Training scientists to use ICT in their research activities creates additional opportunities for scientific and technical progress. This training can be conducted in magistrates, postgraduate and doctorate curricula. In this research we presented the project that is now being realized using the DC in Kherson State University as a case study. The next phase of our research is the investigation of the efficiency of using the described model and its influence on qualitative characteristics of science.

References

1. Savchenko, O.Y.: The Learning Abilities as a Key Competence of Secondary Education Competence Approach in a Modern Education: World Experience and Ukrainian Prospects: Library of Educational Policy, 35--46. K.: «K.I.S.» (2004) (In Ukrainian)
2. David, P. A.: Understanding the emergence of 'open science' institutions: functionalist economics in historical context. Industrial and Corporate Change 13, 571--589 (2004)
3. Bohle, S.: What is E-science and How Should it Be Managed? Nature.com, Spektrum der Wissenschaft, http://www.scilogs.com/scientific_an d_medical_libraries/what-is-e-science-and-how-should-it-be-managed
4. Towards knowledge societies: UNESCO world report, http://unesdoc.unesco.org/images/0014/001418/141843e.pdf
5. IEEE International Conference on eScience, https://escience-conference.org
6. Science under lock. The second part, http://habrahabr.ru/post/190046
7. PLOS is anti-elitist! PLOS is elitist! The weird world of open access journalism, http://www.michaeleisen.org
8. Pagano, U., Rossi, M. A.: The crash of the knowledge economy Camb. J. Econ. 33 (4), 665--683. (2009)
9. Ranking Web or Webometrics, http://www.webometrics.info
10. Gardner, G.G.: On-Line Education: Developing Competitive. Informational Technologies in Education 1, 22--25 (2008)
11. Ermolayev, V.A., Spivakovsky, A.V., Zholtkevych, G.N.: UNIT-NET IEDI: An Infrastructure for Electronic Data Interchange. Informational Technologies in Education 1, 26--42 (2008)
12. Spivakovsky, A., Alferova, L., Alferov, E.: University as a corporation which serves educational interests. In: Ermolayev, V., Mayr, H.C., Nikitchenko, M., Spivakovsky, A., Zholtkevych, G. (eds.) Information and Communication Technologies in Education, Research, and Industrial Applications. ICT in Education, Research and Industrial Applications. CCIS, vol. 347 pp. 60--71. Springer, Heidelberg (2013)
13. Vinnik, M., Lazarenko, Y., Korzh, Y., Tarasich, Y.: Use of Computer Communication Means for Future Software Engineers' Preparing. J. Pedagogical almanac 21, 100--108 (2014) (In Ukrainian)
14. Kravtsov, H. M., Vinnik, M. O., Tarasich, Y. H.: Research of Influence of Quality of Electronic Educational Resources on Quality of Training With Use of Distance Technologies. Informational Technologies in Education 16, 83--94 (2013) (In Ukrainian)
15. Spivakovsky, A. Klymenko, N., Litvinenko, A.: The Problem of Architecture Design in a Context of Partially Known Requirements of Complex Web Based Application "KSU Feedback". Informational Technologies in Education 15, 83--95 (2013)
16. Spivakovsky, A., Vinnik, M., Tarasich, Y.: To the Problem of ICT Management in Higher Educational Institutions. Information Technologies and Learning Tools 39, 99--116. (2014) (In Ukrainian)
17. Spivakovska, E., Osipova, N., Vinnik, M., Tarasich, Y.: Information Competence of University Students in Ukraine: Development Status and Prospects. In: Ermolayev, V., Mayr H. C., Nikitchenko, M., Spivakovsky, A., Zholtkevych, G. (eds.) ICT in Education, Research and Industrial Applications CCIS, vol. 469, pp. 194--216. Springer, Heidelberg (2014)
18. State Statistics Service of Ukraine, http://www.ukrstat.gov.ua
19. Ministry of Education and Science of Ukraine, http://mon.gov.ua
20. World University Rankings, http://www.timeshighereducation.co.uk/world-university-rankings

On the Results of a Study of the Willingness and the Readiness to Use Dynamic Mathematics Software by Future Math Teachers

Olena Semenikhina[1], Marina Drushlyak[1]

[1] Sumy Makarenko State Pedagogical University, Romenska St. 87, Sumy, Ukraine

e.semenikhina@fizmatsspu.sumy.ua, marydru@mail.ru

Abstract. The article presents the results of pedagogical research on the willingness and the psychological readiness to use dynamic mathematics software by future math teachers. We used nonparametric method for dependent samples – the McNemar's test. The hypothesis, that the study of Special course on the use of dynamic mathematics software for future teachers has a positive impact on the willingness and the psychological readiness to use such software in their own professional activities, is confirmed at the significance level of 0.05. Additionally, the results of the experiment on the willingness and the readiness to support the teaching of some subjects (algebra, planimetry, solid geometry and analysis) with dynamic mathematics software and the willingness and the readiness to use some dynamic mathematics software (*Gran (Gran1, Gran2d, Gran3d), GeoGebra, Cabri, MathKit, DG, GS)* by Ukrainian math teachers is given.

Keywords. The study of mathematics, computer applications in the study of mathematics, special course, dynamic mathematics software, the McNemar's test.

Key Terms. InformationCommunicationTechnology, TeachingProcess, TeachingMethodology.

1 Introduction

Ukrainian education has always tried to involve leading technologies and tools that have spread in the world and improve the level of education of ordinary Ukrainians. That is why since the end of the last century information technology has started to be actively implemented in the learning process (also in mathematics). Specialized software appeared and the main purpose of them was computational and visual support of solving of math problems. Later the software, that allows to model processes and to observe the changes in constructions, appeared. But the use of such

software was limited in schools because of a number of reasons, among which the insufficient technical equipment of schools, the lack of targeted preparation of teachers to use specialized software, the lack of software with a clear (Ukrainian, Russian) interface, a small number of teacher's guides, etc.

Now there is a great amount of software which can be used in teaching mathematics. We previously conducted an analisis of the current tendencies of mathematics software use in education in [1]. But the workload of school teachers does not let them to monitor the appearance of such software, to learn the tools and to use them at their lessons. The main part of Ukrainian math teachers are 40 and more years old. This means that they were not acquainted with mathematics software during their preparation, and they used information technologies on the level of Internet users and *Word, Excel, PowerPoint*. They do not use software consciously, because they believe that chalk-and-Board style is better at studing mathematics.

These and other reasons have led us not only to enter the Special course of the study of mathematics software in the curricula of preparation of modern teachers, but to study the impact of this course on the willingness and the readiness to use mathematics software in the professional activity of math teachers.

2 Research of the Willingness and the Psychological Readiness of Future Math Teachers to Use Dynamic Mathematics Software

During 2010-2014, we have investigated the problem of the willingness and the psychological readiness to use mathematics software by future math teachers [2].

We allocate dynamic mathematics software (DMS), that can model and modify mathematical objects interactively. We consider *Gran, DG* (Ukraine), *GeoGebra* (*GG*, Austria), *MathKit*, *Live Mathematics (LM)* (Russia), *Cabri* (France), *The geometer's Sketchpad* (*GS*, USA), etc. We allocate these software for the following reasons: 1) software *Gran* and *DG* are recommended by the Ministry of Education and Science of Ukraine; 2) software *MathKit* and *Live Mathematics* are actively used by Russian teachers, as evidenced by a great number of methodological works of math teachers; 3) software *Cabri*, *The geometer's Sketchpad* and *GeoGebra* are the most popular in the world, as evidenced by the numerous translations of monographs and multi-lingual interfaces of these software.The work with them intuitive and identical – basic objects are built, then they can be dynamically changed and user can observe certain quality properties and quantitative characteristics. The study of features of these software and recommendations for their use are generalized by us in [3-10].

The base of the reaserch was Sumy Makarenko State Pedagogical University. Preparation of math teachers is in accordance with the curricula. The introduction of these software was during the study of methodology of mathematics and during the study of a special course "Computer Applications in the Study of Mathematics" (further Special course). The program of the Special course was described in [11-13] and was improved during the years 2008-2014. The experience of the involvement of

dynamic mathematics software in support of teaching mathematics in the school was during teaching practice (see Table 1).

Table 1. The extract from the curriculum of the speciality "Mathematics*"

Course	Feachers		Note
Methodology of mathematics	Semester	6;7;8	The course contains the module "Computer support for learning mathematics" – 7-th semester, 12 hours.
	Credits	2,5;2;2	
	Class hours	46;46;44	
Teaching practice	Semester	8	At the beginning of the third quarter, within 2 months, on the basis of city schools It is supposed to teach 10 math lessons at 5-9 classes
	Credits	6	
Special course	Semester	8	It is supposed to study different dynamic mathematics software during solving algebra, geometry, analysis problems
	Credits	3,5	
	Class hours	50	

At the beginning of the teaching practice students learn how to solve mathematical problem with the use of dynamic mathematics software (DMS) at the lessons of Special courses. During the teaching practice they have the opportunity to see (or not to see) and analyze lessons of those teachers who use DMS in their own professional activity.

We believe that during this period the basis for the motivation of the learning and further use of DMS in professional activities is formed. Therefore, the Special course, which is studied immediately after the teaching practice, becomes the factor of impact on the student, which gives the opportunity to talk about the willingness and the readiness to use DMS in the future professional activity.

Because these personal characteristics can be formed within the teaching of the Special course, it was natural to involve such statistical methods, that give the opportunity to talk about the dynamics of change based on data about the initial and final state of the object. So we fixed the internal state of the willingness and the psychological readiness of the student to use DMS with the help of questionnaires at the beginning and at the end of the study of the Special course (see Table 2).

Table 2. The questionnaire

№	Questionnaire	Answers
1.	Do You need to use DMS at the lessons of algebra (planimetry, solid geometry, mathematical analysis)? Why?	Yes Yes, not at all No

№	Questionnaire	Answers
2.	Do You want to use DMS at the lessons of: a) algebra; b) planimetry; c) solid geometry; g) mathematical analysis? Why?	 Yes/No Yes/No Yes/No Yes/No
3.	Do You feel readiness to use DMS at the lessons of: a) algebra; b) planimetry; c) solid geometry; g) mathematical analysis? Why?	 Yes/No Yes/No Yes/No Yes/No
4.	Specify a priority of DMS that You like. Specify a priority of DMS, which is better to use at math lessons on Your opinion.	*Gran* *DG* *GG* *MathKit* *GS* *Cabri*

It was applied the McNemar's test [14], because the scale of results in questions 1-3 has two items ("Yes" or "No"). This method is nonparametric and used to compare distributions of objects in two sets according to some property on the scale with two categories (e.g., "like - dislike ", "ready - not ready," "willing - unwilling" and others).

For a McNemar's test the following conditions are required: 1) random sample; 2) dependent sample; 3) pairs (x_i, y_i) are mutually independent (the members of the sample have no effect on each other); 4) the scale has only two categories.

The research was conducted from 2010 to 2014. Each year we have accumulated the results of the sample with volume 37, 35, 38, 37, 31, respectively. The total number of respondents amounted to 178 people. We selected results from them at random.

2.1 The Use of Dynamic Mathematics Software in the Study of Mathematics in Secondary Schools

The beginning of our research was associated with the study of the status of the use of DMS in the study of mathematics in secondary schools. Through conversations with teachers, graduates, teachers-methodists of our region it was found that the "poor" use of DMS in the learning process is not only due to the limited number of computers in schools, but due to lack of the willingness of teachers to involve such software to the solution of mathematical problems. Although they did not deny the feasibility of this approach, but noted, among other things, about the inability to use DMS (68%), the need for additional time to study them (87%), the small number of methodological literature on the use of DMS (90%) and the lack of collections of tasks, which can be solved by using DMS (36%).

2.2 The research of the Willingness of Future Math Teachers to Use Dynamic Mathematics Software in Their Professional Activities

Searching for ways to solve the problem, we have suggested that a focused study of the Special course will have a positive impact on the willingness of future math teachers to use DMS in their profession.

The test of the assumption was carried out according the McNemar's test on taken results in the amount of 30 pieces from 178 at random.

Hypothesis H_0: the Special course does not impact on the willingness of students to use DMS in the future math teacher's profession. Hypothesis H_a: the Special course has a positive impact on the willingness of future math teachers to use DMS.

We had two series of observations: $X=\{x_1, x_2, ..., x_N\}$ and $Y=\{y_1, y_2, ..., y_N\}$, where (x_i, y_i) are the results of measuring of the willingness to use DMS in the future professional activity of the same object (the willingness of the student before and after the Special course). In our notation, x_i or y_i takes the value 0 if the object of study does not wish to use DMS at any of the classes (algebra, planimetry, analysis, solid geometry) and 1 otherwise. The results of the dual survey recorded in the Table 3.

Table 3. The survey on the willingness to use DMS

	The second surve		
The first surve	$y_i=0$	$y_i=1$	
$x_i=0$	a=6	b=10	a+b=16
$x_i=1$	c=2	d=12	c+d=14
	a+c=8	b+d=22	N=30

In the conditions of the experiment parameter a determined the number of students who both times said "No"; the parameter b was the number of students who the first time said "No" and the second time said "Yes"; the parameter c was the number of students who the first time said "Yes" and the second time said "No"; the parameter d was the number of students who both times said "Yes".

To apply the McNemar's test we will find $T_{exper}= min(b,c)$, if $n = b + c < 21$. For our data $T_{exper}= 2$, since $n = 10 + 2 = 12 < 20$. Statistics of the criterion for significance level $\alpha = 0,05$ is $p = 0,019$. According to the rule of decision [14] we have $0,019 < 0.025$. We have to reject hypothesis H_0 and accept the alternative one, and since $b > c$, then we consider that the impact of the study of the Special course on the willingness to use DMS is not only statistically correct, but also positive.

2.3 The research of the Readiness of Future Math Teachers to Use Dynamic Mathematics Software in Their Professional Activities

In parallel with the research of the willingness to use DMS we explore the personal readiness of future math teachers to use DMS in their professional activities (question 3 of the questionnaire).

The hypothesis H_0: the Special course does not impact on the psychological readiness of students to use DMS in their professional activities.

Then the hypothesis H_a: the Special course impacts on the psychological readiness of future math teachers to use DMS.

The test of the assumption was carried out according the McNemar's test on taken results in 40 pieces from 178 questionnaires at random (see Table 4).

Table 4. The survey of psychological readiness to use DMS

	The second surve		
The first surve	y_i=0	y_i=1	
x_i=0	a=7	b=16	a+b=23
x_i=1	c=6	d=11	c+d=17
	a+c=13	b+d=27	N=40

Since $n = b + c = 22 > 20$, the statistics of criterion is calculated according the formula $T_{exper}= (b - c)^2 / (b + c) = 4,54$. The assumption of the fairness of the null hypothesis is approximated like the χ^2 distribution with one degree of freedom ($\upsilon=1$). For significance level $\alpha=0,05$ the critical value of the test is $T_{critic}=3,84$. The obtained value of $T_{exper}=4,54 > T_{critic}=3,84$, therefore, the hypothesis H_0 is rejected and the alternative hypothesis, indicating that the impact of the Special course on the readiness to use DMS in future professional activity is significant and cannot be explained by random variation, is accepted.

2.4 The research of the Willingness of Future Math Teachers to Use Different Dynamic Mathematics Software in Their Professional Activities in Teaching of Some Subjects

Because the questionnaire was on the research of the willingness to use DMS at the lessons of algebra, planimetry, solid geometry and analysis, and on the research of the use of different DMS (*Gran (Gran1, Gran2d, Gran3d), GeoGebra, Cabri, MathKit, DG, GS*), we were able to fix and process results about the willingness to use DMS in teaching of some subjects – algebra, planimetry, solid geometry, analysis (see Table 5) and about the willingness to use different DMS – *Gran, DG, GeoGebra, MathKit, GS, Cabri* (see Table 6).

For each position of the table 5 we have the rejection of the null hypothesis H_0 and the acceptance of alternative hypothesis, i.e., at the significance level $\alpha=0,05$ we can say about the positive impact of the studing of the Special course on the willingness of future math teachers to use DMS at the lessons of algebra, planimetry, analysis and solid geometry.

Table 5. The survey of the willingness to use DMS in teaching of different subjects

Question. Do You wish to use DMS at the lessons of:	Quantitative indices					Indices of the McNemar's test (α=0,05)				
	a	b	c	d	N	b+c	$T_{ек}$	P	H_o	H_a
algebra	6	11	2	11	30	13	2	0,011	0	1
planimetry	2	15	5	8	30	20	5	0,021	0	1
analysis	5	12	3	10	30	15	3	0,018	0	1
solid geometry	6	14	4	6	30	18	4	0,015	0	1

Table 6. The survey of the willingness to use some DMS

Question. Do You wish to use:	Quantitative indices					Indices of the McNemar's test (α=0,05)				
	a	b	c	d	N	b+c	$T_{ек}$	P	H_o	H_a
Gran	8	11	2	9	30	13	2	0,011	0	1
DG	5	12	3	10	30	15	3	0,018	0	1
GG	2	12	2	14	30	14	2	0,006	0	1
MathKit	6	14	4	6	30	18	4	0,015	0	1
GS	12	10	6	2	30	16	6	0,227	1	0
Cabri	20	6	3	1	30	9	3	0,254	1	0

For indeces of the table 6 we have the acceptance of hypothesis H_0 for the last two rows. This means that at the significance level 0.05, future math teachers wish to use software *Gran, DG, GG, MathKit*, but we have no reason to say about the willingness to use *GS* and *Cabri*. We can explain this because of "poor" interface of *GS* and the absence of Ukrainian (or Russian) interface of *Cabri.*

Visualization of the obtained results during the experiment years is given in Fig. 1-2.

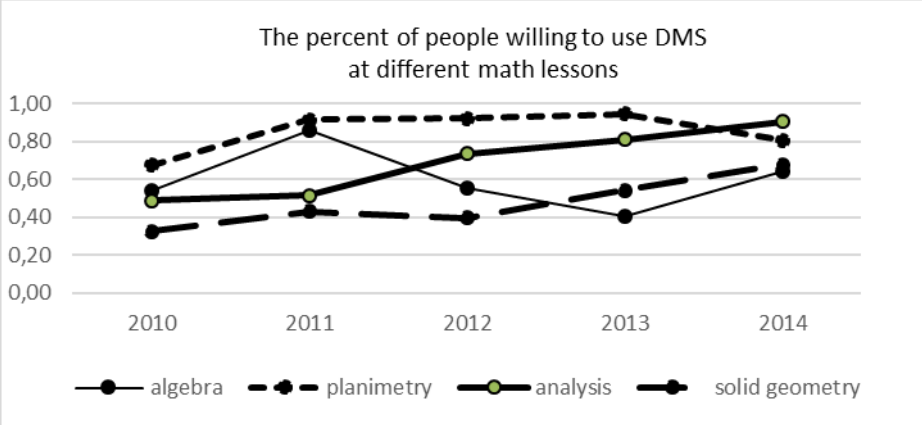


Fig. 1. The percent of people willing to use DMS at different math lessons

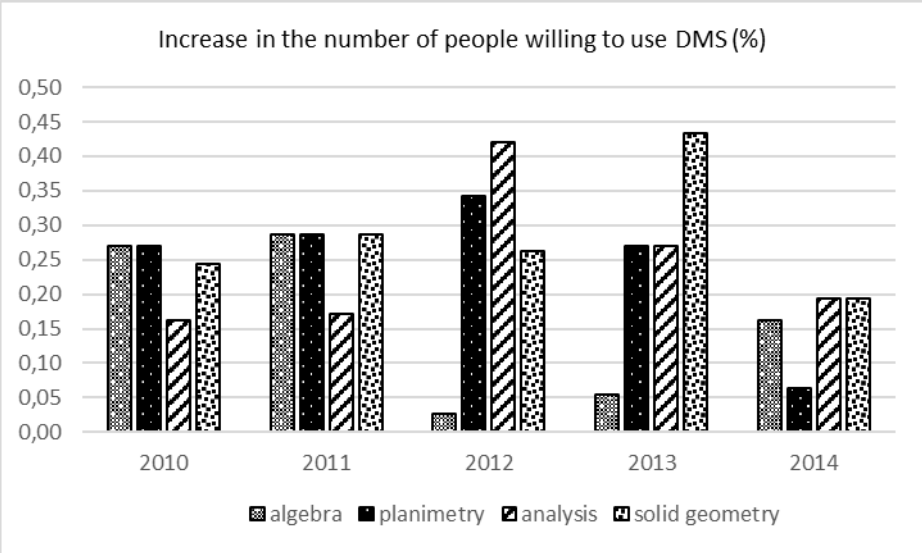


Fig. 2. Increase in the number of people willing to use DMS (%)

Also we give some information about the "attractiveness" of software according to the survey of future and working math teachers, which was conducted at scientific-methodical seminars (on the basis of physics and mathematics faculty) (see Table 7, Fig. 3-10).

Table 7. The attractiveness of software (%)

	Gran		DG		GG	
Year	**T**	**S**	**T**	**S**	**T**	**S**
2010	0,93	0,59	0,74	0,68	0,28	0,68
2011	0,75	0,71	0,51	0,80	0,32	0,91
2012	0,86	0,71	0,83	0,66	0,45	0,79
2013	0,68	0,43	0,54	0,54	0,68	0,78
2014	0,40	0,32	0,13	0,48	0,66	0,97
	MathKit		**GS**		**Cabri**	
Year	**T**	**S**	**T**	**S**	**T**	**S**
2010	0,11	0,32	0,00	0,24	0,00	0,00
2011	0,11	0,57	0,02	0,43	0,00	0,00
2012	0,17	0,66	0,08	0,32	0,00	0,11
2013	0,19	0,86	0,03	0,35	0,05	0,08
2014	0,13	0,94	0,00	0,19	0,07	0,13

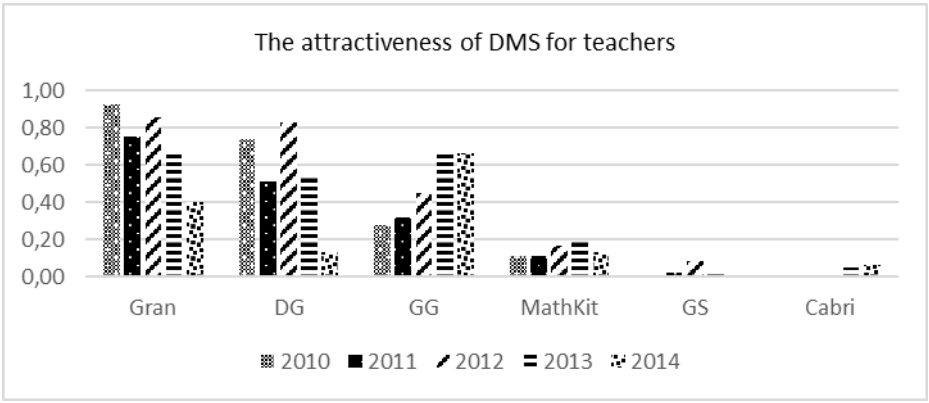


Fig. 3. The attractiveness of DMS for teachers

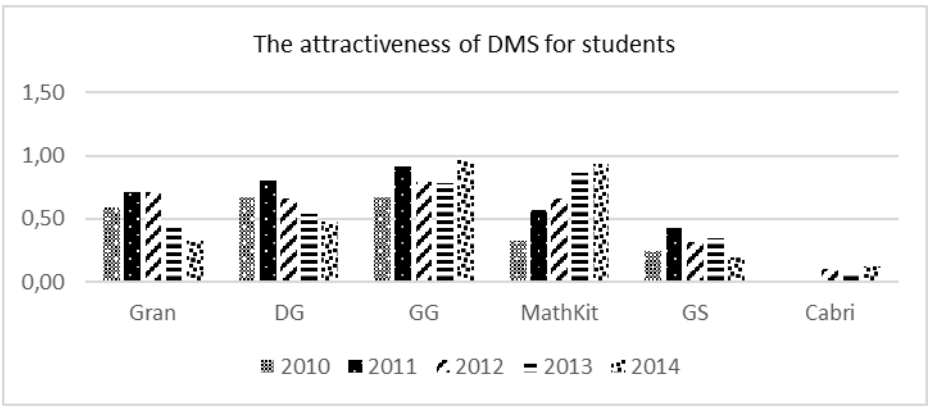


Fig. 4. The attractiveness of DMS for students

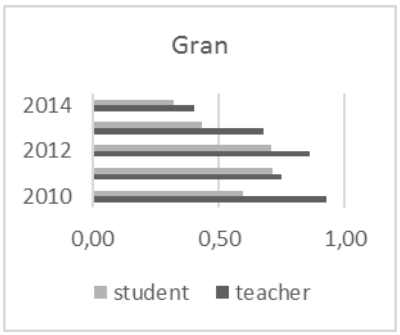


Fig. 5. The attractiveness of *GRAN*

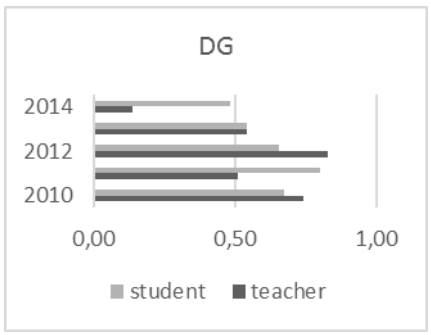


Fig. 6. The attractiveness of *DG*

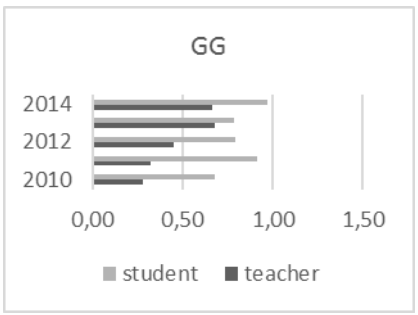


Fig.7. The attractiveness of *GG*

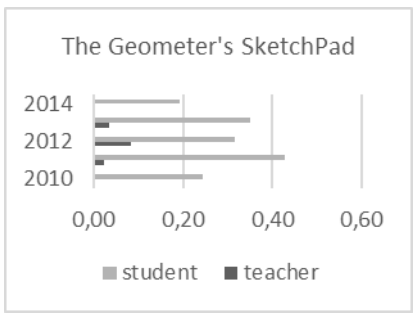
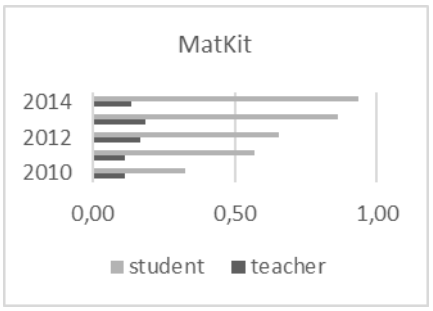


Fig.8. The attractiveness of *GS*



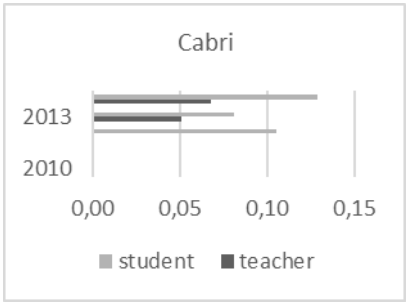


Fig. 9. The attractiveness of *MathKit*

Fig. 10. The attractiveness of *Cabri*

2.5 The research of the Readiness of Future Math Teachers to Use Different Dynamic Mathematics Software in Their Professional Activities in Teaching of Some Subjects

Because the questionnaire was on the research of the psychological readiness to use DMS at the lessons of algebra, planimetry, solid geometry and analysis, as well as the readiness to use different DMS (*Gran (Gran1, Gran2d, Gran3d), GeoGebra, Cabri, MathKit, DG, GS*), then we could fix the results of the readiness to use DMS in teaching of different subjects (algebra, planimetry, solid geometry, analysis) (see Table 8).

Table 8. The survey of the readiness to use DMS in teaching of different subjects

Do You feel the readiness to use DMS at the lessons of:	**Quantitative indices**					**Indices of the McNemar's test (α=0,05)**			
	a	**b**	**c**	**d**	**N**	**n=b+c**	**T_2**	**H_0**	**H_a**
algebra	6	17	7	10	40	24	4,17	0	1
planimetry	2	21	9	8	40	30	4,80	0	1
analysis	5	18	7	10	40	25	4,84	0	1
solid geometry	4	17	15	4	40	32	0,13	1	0

For all items, except the last, we have the rejection of the null hypothesis H_0 and the acceptance of the alternative hypothesis, i.e., at the significance level $\alpha=0,05$, we can say about the positive impact of studing of the Special course on the psychological readiness of future math teachers to use DMS at the lessons of algebra, planimetry, analysis. However, experimental results do not give grounds to say about the positive impact on the readiness to use DMS at the lessons of solid geometry. Increase in the number of students who feel the readiness to use DMS at the math lessons is presented in Fig. 11.

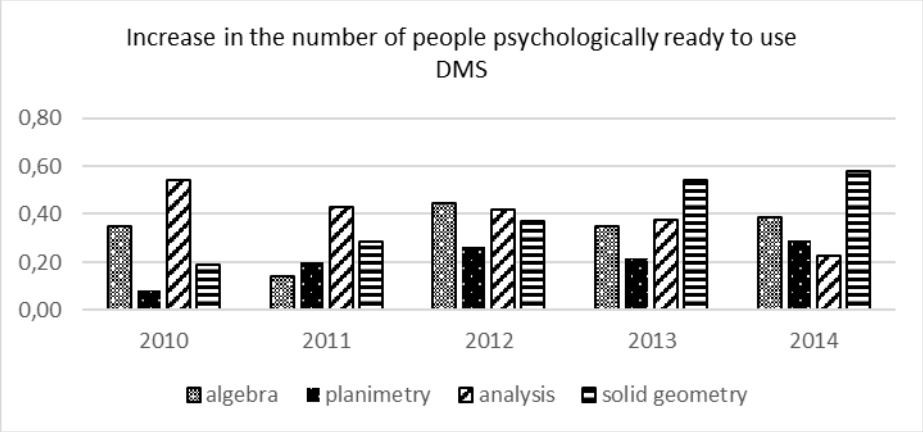


Fig. 11. Increase in the number of people psychologically ready to use DMS

3 Conclusion

Thus, this research allows to state the following.

1. Future math teachers understand the need to use DMS and welcome the studing of the Special course, since the research of the willingness and the readiness to use DMS demonstrates a positive dynamics. The assumption about the positive impact of the studing of the Special course on the psychological state of students is confirmed at the significance level of 0.05 according to the McNemar's test. In other words, after the studying of the Special course "Computer Applications in Teaching Mathematics" the number of students, who have the willingness and feel the readiness to use DMS in future professional activity, increases.

2. Most students focused on using DMS at the lessons of algebra, planimetry and analysis. We explain this because of not only a sufficient number of DMS and good tools in such software, but enough number of teacher's guide for their application and free access to DMS with Ukrainian or Russian interface.

The percentage of students, who are willing to use DMS at the lessons of solid geometry, is too small. We explain this not only because of small number of software and "poor" tools in such software, but of lack of Russian or Ukrainian interface in them. Also there is the lack of methodical material of solving solid geometry problems with the use of specialized software.

3. Teachers, who work at the school, have the willingness and the inner readiness to use DMS, but face with the limited access to computer classes. The involvement of DMS, as they say, is possible only during extracurricular activities.

4. *GRAN* and *GeoGebra* are the most popular in Ukraine. In recent years there has been a decline in the use of the first and great attachment to the second. We explain that because of free access and frequent updating of *GeoGebra*, the steady growth of

its tools (in particular, the version *GeoGebra 5.0* with 3d-tools was tested in 2013, and is distributed now).

5. Russian software *MathKit* finds his supporters (the latest version is license, but the early versions can be found in Internet). It is attractive because of ”rich” tools and automated control, which is not provided in other DMS.

6. Students and teachers prefer *Gran* and *DG*. We explain that because of the free distribution, the Ukrainian interface, a sufficient number of researches in periodicals, the recommendations of the Ministry of Education and Science of Ukraine (also at the lessons of computer science).

Note that students prefer *GS*, *MathKit* and *Cabri* more then teachers. We explain that becourse of the lack of Ukrainian interface, the license and the unwillingness of teachers to work with unfamiliar DMS.

7. According to the research we note the increasing demand for *GeoGebra* (it was pointed out by the future and working math teachers). We believe that it is necessary to pay attention just at it, because *GG* is continuously updated, freely distributed, has interface on 30 languages, that confirms its popularity.

8. Future research should be conducted towards the creation of methodical support of school math courses based on *GG*. During the preparation of future math teachers we need to focus not only on traditional for the Ukrainian school software *Gran, DG*, but also on the other DMS, which are widely distributed in Internet and used by teachers.

References

1. Semenikhina, E., Drushlyak, M.: The Necessity to Reform the Mathtmatics Education in the Ukraine. Journal of Research in Innovative Teaching. 8, 51--62 (2015)
2. Semenikhina, O., Shishenko, I.: Consequences of the Spread of IT and the Shift in Emphasis of Teaching Mathematics in Higher School. Vyscha Osvita Ukrainy. 4, 71--79 (2013)
3. Semenikhina, O., Drushlyak, M.: Use of Computer Toos of IGE CABRY 3D at a Solving of Stereometry Problems. Informatika ta informatsiyni tehnologiyi v navchalnih zakladah. 4, 36--41 (2014)
4. Semenikhina, O., Drushlyak, M.: Computer Tools of Dynamic Mathematics Software and Methodical Problems of Their Use. Information Technologies and Learning Tools. 42 (4), 109--117, http://journal.iitta.gov.ua/index.php/itlt/article/view/1055#.VCqAD0Hj5nE. (2014)
5. Semenikhina, O., Drushlyak, M.: On Checking Tools in the IGE MathKit. Naukoviy visnik Melitopilskogo derzhavnogo pedagogichnogo universitetu. Seriya: Pedagogika. 13 (2), 189--195 (2014)

6. Semenikhina, O., Drushlyak, M.: Geometric Transformations of the Plane and Computer Tools for Their Implementation. Komp'yuter v shkoli i sim'yi. 7(119), 25--29 (2014)
7. Semenikhina, E., Drushlyak, M.: Computer Mathematical Tools: Practical Experience of Learning to Use Them. European Journal of Contemporary Education. 9 (3), 175--183 (2014)
8. Drushlyak, M.: Computer Tools "Trace" and "Locus" in Dynamic Mathematics Software. European Journal of Contemporary Education. 10 (4), 204--214 (2014)
9. Semenikhina, E., Drushlyak, M.: Creation of Own Computer Tools in the Dynamic Mathematics Environment. Informatika ta informatsiyni tehnologiyi v navchalnih zakladah. 5(53), 60--69 (2014)
10. Semenikhina, E., Drushlyak, M.: GeoGebra 5.0 Tools and Their Use in Solving Solid Geometry Problems. Information Technologies and Learning Tools. 44(6), 124--133, http://journal.iitta.gov.ua/index.php/itlt/article/view/1138/866#.VKKRJc-eABM (2014)
11. Semenikhina, E.: The Course for the Study of Dynamic Mathematics Software as a Necessary Component of Training of Modern Math Teacher. In: Modern Trends in Physics and Mathematics Education: School – University. Revised Extended Papers of International scientific-practical conference, pp. 75--78, Solikamsk State Pedagogical Institute (2014)
12. Semenikhina, E.: On the Nessesity to Introduce Special Courses on Computer Mathematics. Vestnik TulGU. Seriya. Sovremennyie obrazovatelnyie tehnologii v prepodavanii estestvenno-nauchnih distsiplin, 12, 102--107 (2013)
13. Semenikhina, O., Drushlyak, M.: The Study of Specialized Mathematics Software in the Context of the Development of the System of Math Teachers Preparation. In: Proceedings of IX International Conference ITEA-2014, pp. 61--66, International Research and Training Center for Information Technologies and Systems, Kyiv (2014)
14. Grabar, M., Krasnyanskaya, K.: The Application of Mathematical Statistics in Educational Research. Nonparametric Methods. Pedagogika, Moscow (1977)

An Analysis of Video Lecture in MOOC

Jyoti Chauhan[1], and Anita Goel[2]
[1] Department of Computer Science ,University of Delhi, Delhi, India
jyotich2009@gmail.com
[2] Department of Computer Science, Dyal Singh College, University of Delhi, Lodhi road, New Delhi, India
goel.anita@gmail.com

Abstract. Video is a content delivery form used for delivering lecture content in Massive Open Online Course (MOOC). While institutions plan to launch MOOC on their own platform or adapt an existing one, there is a need to specify the features required for video lecture in MOOC. In this paper, we present a checklist of features for video lecture incorporated in MOOC from the learner's perspective. The use case based approach has been followed for identifying the features of video lecture in MOOC. The checklist helps during requirement specification of video in MOOC as the provider select the desired features from the checklist.

Keywords. Video Lecture, Video Analysis, MOOC, Online Education, Feature Checklist

KeyTerms. ICT Component, Characteristic, Academia, Environment, Management

1 Introduction

Online learning uses technology and electronic media for delivering and receiving courses. It is considered as most promising development in education that provides education with technology. With technology globalization, the concept and methodology of learning, and teaching has undergone a change. The technology usage in education provides global learning environment that allows accessing the course material anytime, anywhere, connect other students, and get access to the content without considering any geographical boundary. The significant changes in technology usage in online education has seen emergence of MOOC in 2008. It is a popular way to offer online courses globally by the universities and education providers.

MOOC uses web-based tools and environments to deliver education [1]. It provides online courses aimed at unlimited participation and open access via the web [2]. It is being used across the globe for offering online courses. Some of the popular MOOC providers are - Coursera[1], edX[2] and Udacity[3] in United States, FutureLearn[4] in

[1] https://www.coursera.org
[2] https://www.edx.org/
[3] www.udacity.com
[4] www.futurelearn.com

United Kingdom, iversity[5] in Germany, FUN[6] in France and MiríadaX[7] in Spain [3]. The course lectures in MOOC are delivered in different formats, like, text books, lecture slides, academic papers, tutorial notes, video lectures, blog posts, article links, quizzes and assignments.

In MOOC, video is a primary delivery mechanism to publish the recorded lecture content. It consists of audio (voice of instructor), visual (video of lecture) and text (caption/transcript/subtitles) in one package. The study of edX [4, 5] analyzed that students spend most of their time on video lectures. With the increasing popularity of the video lecture, MOOC providers are continuously improving video lecture content delivery as well as its production.

A video lecture in MOOC is a combination of several elements like, lecture by the teacher, quiz, and lecture slides. For viewing the video lecture, MOOC provides an interface to the student. The video interface has several controls, using which the student can perform settings for the view of the video lecture. The content in video lecture is delivered in different types and formats. The accessing of the content is possible with video lecture options.

For offering MOOC, interested institutions have an option to go for self-hosted platforms or use proprietary platform. When using self-hosted platform, the providers have a choice to 1) develop their own MOOC platform, or 2) use open source platform. Generally, the open source platform is a preferred choice, which may require modification and customization as per the user needs. When developing a new MOOC platform, there is a need to specify the features that has to be provided to the video lectures presented here. Although video lecture are being delivered by MOOC providers, there is no mention of its feature specifications.

In this paper, we focus on creation of the requirement checklist for the video in MOOC. It helps during the development, in selecting and specifying the requirements for video lecture to be included in MOOC.

Here, we present the feature requirement of video lecture that facilitates selecting requirements for functionalities of video in MOOC. The functionality of MOOC video is classified in two components, namely, 1) *Video Interface*, and 2) *Video Lecture Content*. It aids the MOOC platform developer during requirement specification phase. The features required for MOOC video can be selected from checklist.

For formulating the feature checklist of video lecture in MOOC, a study of MOOC platforms was conducted. We chose popular MOOC providers that provide different functionalities for delivered video lecture. We studied Coursera, edX, and Udacity. “Coursera is by far the largest MOOC provider” [6] reported by Class Central. The most popular MOOC providers as reports by [7] [8] [9] are Coursera, edX, and Udacity. Since these are using different platforms, covers diversity in terms of functionalities provided and mechanism used to deliver video content, our analysis can be applied to any video lecture in a MOOC.

The video feature requirement presented has been used to a few open source platforms to identify the functionality provided by them. The checklist of feature has been applied for the functionality visible to student.

In this paper, Section 2 is a survey of related work. Section 3 provides an overview of the video lectures in MOOC. Section 4 explains the methodology used for our study. Section 5 discusses the controls of video interface. Section 6 explains the video lecture

[5] https://iversity.org/
[6] https://www.france-universite-numerique-mooc.fr/
[7] https://www.miriadax.net/

content. Section 7 describes the analysis of video lecture feature in detail. Section 8 illustrates some examples on which our analyzed features have been applied. Section 9 lists the benefits. Section 10 enumerates the limitation. Section 11 states the conclusion.

2 Related Work

For video in MOOC, much work has been done related to understanding student behavior. In [4] [10] [11] [12] [13], student behavior is studied in quantitative way. The analysis using edX [4] and Coursera course [10] focus on student engagement using action time of students. Student engagement studies using different methods, for example, using clickstream metrics [11], navigation study [12] and use of framework [13]. The qualitative findings of clickstream are combined with cognitive science [11] and student's goals [10]. These studies have not focused on video lecture particularly.

Some authors [14] [15] [16] focuses on different aspects of videos, like, the video interface, its features and properties. Guo et al. [14] studies the student behavioral on edX platform using different properties of videos like, length, speaking rate, video type, and production type. Guidelines for the video lecture are presented by Chorianopoulos et al. [15] with the focus on video style, editing, sharing, controlling and analytics. In [16] Kim et al. define design implication guidelines for video interfaces based on the student engagement, like, providing shorter videos and navigation links.

Ortega et al. [17] perform a study based on different ECO MOOC platforms OpenMOOC, Open EdX, iMOOC etc. and external MOOC platforms - Coursera, Udacity, MiriadaX, OpenCourseWare-MIT, Futurelearn, and iversity. Their focus is to study "accessibility" of MOOC platform including the video lecture. The recommendations are about subtitles (vocal or non-vocal sounds), secondary screen integration, downloadable text, possibility of text reader processing, interface navigation by keyboard. However, this study is based on the published literature only and does not cover the different aspects of video lecture.

3 Video Lectures in MOOC

Video lectures are the pre-recorded learning material that acts as a medium of communication between the student and lecture. In MOOC, the video lectures are primarily used to deliver lecture content. MOOC video lectures are considered as central to the student learning experience [14]. The courses are structured as list of video lectures consisting of activities and contents, like, walkthroughs, assessment problems, quizzes etc. Students need to watch the video lectures of the courses for learning. Video provides self-regulated and independent learning. It has transformed the traditional classrooms by replacing "one-size-fits-all" approach with self-paced learning, and from curriculum/teacher centric to student centric learning.

Video in MOOC provides an interface to the students for managing the delivered lecture. The interface provides different kinds of controls to the students. These controls help to, navigate and view the content by play, pause, stop, increase/decrease speed, volume and toggling to full screen mode etc. It also allows the student to

download the video in different formats and view it offline. The video lectures may also be made available to students who are not enrolled in the course, via YouTube[8] and other video sharing websites.

The video lecture content is the lecture delivered by the teacher which may contain caption and in-video activity like, quiz. The lecture material provided with a video lecture may be presentation slides, transcript of video, related document etc. The lecture material is provided in different formats. According to Clark and Mayer [18], the transcript helps in understanding the complex domain-specific video lecture content. Sometimes the text content of video is presented in different languages. Different MOOC providers use different types of video presentation styles, lecture material, video interfaces and activities for the video lecture.

4 Methodology

MOOC providers use diverse mechanisms to deliver their video lecture. Videos are different in terms of the available features, delivered content and formats and interface used to display the lecture video. The features not only vary with the provider but even with change in MOOC with the same provider. We gathered the information about video by viewing the video lectures by different MOOC providers.

For our study, we selected three most popular MOOC providers - Coursera, edX and Udacity. For our sample data set, we identified courses provided by various universities in different areas, for diversity. The courses provided by universities, like, Stanford, Duke and Harvard; in subjects like, medical chemistry, biology and computer science, were selected. We watched more than two thousand videos across the three different providers and manually analyzed them. Our sample set includes videos of short and long duration, interactive and non-interactive with different presentation styles. Table 1 shows MOOC providers, courses offered, university offering the course and number of videos viewed, from the selected sample.

Our experience and observation while watching the video lecture act as the baseline for the extraction of features and their segregation into different categories.

The long duration videos are time consuming to watch. So it is required to skip some portions of video but not to miss any important feature. For this reason, the frame of the video must be seen at any instant of time to check the relevance of that frame. This kind of provision is provided by the feature named as *Poster Frame*. But it is not provided by all platforms. As a result, the video need to be watched either by simply *play* or by fast forwarding the video using *speed* +. During playing a video lecture in edX, it showed an error once. Since edX provides *help support* we reported the problem and got it resolved. The reason behind this kind of problem is, at times the video player is not supported by the all browser. Therefore, we may need to change the player which is possible using *change player* feature as provided by Coursera. The language barrier is also a problem noticed in a few courses. According to nationality of different instructors, accent and vocalization of the lecture instruction varies. So the availability of *caption* on the video interface panel is very helpful for better understanding of the content. Also, the caption provides *multilingual support*.

[8] www.youtube.com

Table 1. Video selection for our study.

MOOC Providers	Course	University	Video Watched
Coursera	Compilers	Stanford University	96
	Machine Learning	Stanford University	113
	Introduction to Mathematical Thinking	Stanford University	76
	Child Nutrition and Cooking 2.0	Stanford University	46
	Medical Neuroscience	Duke University	199
	Cryptography I	Stanford University	66
edX	Introduction to Computer science	Harvard University	200
	United States Health Policy	Harvard University	119
	Data Analysis For Genomics	Harvard University	141
	The Chemistry of Life	KyotoUx	120
	Biomedical Imaging	Udx	51
Udacity	Artificial Intelligence In Robotics	Georgia Tech	208
	Exploratory Data Analysis	Facebook	180
	Intro to Computer Science	Univ. of Virginia	312
	Mobile Web Development	Google	176

The availability of *downloading* feature which is provided for offline learning support makes it possible to watch the lecture without dependency on the network connection. Accessibility of *different quality of the video* lecture and *High Definition* (*HD) support* allowed us to watch the video on larger screen with better quality. After watching the video, attempting the quiz and assignment of the course, the student may require to track the progress, like, how many quizzes are there in a lecture, how many them have not been visited yet. But for doing so, the student needs to go back to the lecture and check each lecture separately. The options available are also not easily navigable. Currently the *progress bar* feature is not being provided even by several popular providers like, Coursera. One of the new features noticed in this analysis process is *embedded quiz*. It allows attempting the quiz while watching the video lecture and interacts with the video interface. It helps to check grasping of the learner from the watched video lecture. This experience helped to identify and categorize the features that were analyzed during the video watching sessions.

We analyzed video lecture by the sample data set from perspective of student. It helps to categorize the video lecture into different component and to segregate the features provided by them.

5 Video Interface

MOOC provides video interface to the student for viewing the lecture videos. The interface has several controls that allow students to make settings for the display of the video lecture. We categorize the controls on the video interface into four categories – (1) Display time, (2) View setting, (3) Advance setting, and (4) Help support.

1) ***Display time*** shows the current time during video watching and total duration of the video.
2) ***View setting*** is related to playing the video, like, play/pause, change volume or speed of video. By default, the video is displayed in normal mode which can be toggle to full screen mode. View settings allow the student to control the setting for play (▶), pause (▮▮), speed (+/-), volume (+/-), full screen mode (⛶), navigate to previous/next videos (⏮,⏭) etc. It also allows setting for rewind and replay (↺) video.
3) ***Advance setting*** include are additional controls provided on the video interface. It allows the student to switch to *HD* mode to watch video lecture with better quality. The *poster frame*s of the video on time bar seen at any time without playing it are part of advance setting. The settings also provide multiple *video player* facility which allows changing the player. The video text is provided in different languages that can be chosen from the *caption option*. The format of the caption can also be changed like, font and background color, opacity of window. The controls for advance setting may or may not be provided in all video interfaces.
4) ***Help Support*** allows reporting of problem faced during the watching of video lecture. *Discussion forums* are provided for problem reporting. Help for shortcut keys allow interface navigation using the keyboard.

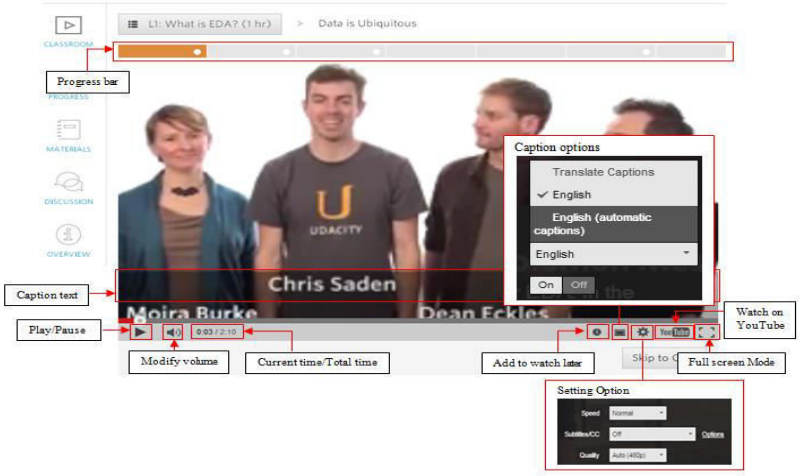


Fig. 1. Video interface controls for video lecture in Udacity.

Different MOOC providers provide different kinds of controls on their video interface. Display setting and view settings are among the common controls provided on video interface. Fig. 1 shows different controls available in the video interface of a video lecture in Udacity - progress bar, caption text, play/pause, volume slider, current time/total time of video, full screen mode, add to watch later, watch on YouTube, caption on/off and language selection, and setting for selecting speed, video quality, and subtitle options.

6 Video Lecture Content

The video lecture content is the lecture delivered by the teacher which may contain caption and in-video activity like, quiz. The support material provided with a video lecture may be presentation slides, transcript of video, related document etc. The support material of the lecture is provided in different formats. The transcript helps in understanding as it provides text of the instruction given by video lecture instructor. Sometimes the text content of video is presented in different languages. Different MOOC providers use different types of video presentation styles, support material, video interfaces and activities for the video lecture.

The lecture content of the video has some activities embedded in it for better explanation and interaction with the students. The video lectures also have accompanying material that helps in better understanding of the lecture. We categorize the content of video lecture into two parts- 1) Lecture Material and 2) Embedded Quiz.

1) ***Lecture Material-*** The video lecture content is provided in the form of slides, video of the lecture, transcript of the video etc. The content is available in different formats, for example, *mp3/mp4, ppt* and *pdf*. The lecture material can be downloaded to support offline learning which may also be available in different qualities.

2) ***Embedded Quiz*** is the in-video activity that is incorporated to make video lecture session more interactive. Students can attempt the quiz while watching the video lecture. The quiz has several options and parameters, like, question type that may be Multiple Choice Question (MCQ), true-false or descriptive.

Different types of video lecture content are delivered by the MOOC providers which have diverse features and controls.

7 Video Lecture Feature Checklist

The features provided by each MOOC platform vary. Therefore to understand the features of the video lecture in MOOC, the analysis has been performed on video lecture of three popular MOOC providers - *Coursera*, *edX* and *Udacity*. In our study the main focus is to identify the features of the video lecture in MOOC. The video lecture in MOOC consists of two components - *Video interface* and *Video Lecture Content*. We arrive at the detailed feature checklist by identifying the features provided by each component of video lecture. We analyzed the features for video lecture components including their *presentation style*.

Video Interface provides controls to the student for viewing the video lecture in MOOC. The features provided by video interface are mainly controls that provide control of the interface to a student. A video lecture also provided different types of content to their student. *Video lecture content* provided several options for accessing the lecture material. The categorization of video interface and video lecture content helps to segregate the provided controls, features and presentation style, from student's perspective.

For arriving at the feature checklist, the features provided by the component and control identified in previous sections are applied on the chosen MOOC providers of our study. From our analysis we find some of the features are provided by each of the MOOC providers, like, current/total time, play/pause, and video mp4 format support. It suggest that these are the basic features need to be made available for student. The inclusion of some feature provides better functioning but does not affect the basic functioning of the video lecture. For example, availability of HD support, multiple languages for caption, availability of different quality video named as optional or advance features etc. These are named as optional or advanced features.

The different levels of availability of feature or control from student's perspective, are classified into three catagories-

- *Basic* – are the most important features that are available in all platforms. It is denoted by weight '3'.
- *Optional* – are the features that are not necessary but may be helpful. The feature supported by any of the two platforms is an optional feature. It is denoted by weight '2'.
- *Advanced* – are the features that are required for specific purpose. Feature available only by single platform is an advanced feature. It is denoted by weight '1'.

On the basis of availibility of a feature, weights are assighned to each identified feature. The weighted feature checklist of features helps to select the features and options that need to be made available to a student in video lecture of MOOC. Table 2 lists the weighted feature checklist of video lecture in MOOC. Each feature is presented in different way by the MOOC providers.

The *Video Interface* and its options are provided in diverse styles. It is displayed either as a separate pop-up window or incorporated in the lecture page. Each provider use different presentation style for *viewing controls*, like, video speed is displayed in terms of range of pixels from .50x to 2.0x. Navigation of the video lecture is control using previous/next control, selecting from the video lecture list or from the progress bar that displays the video lectures. An advance setting control, to change format of the caption allow changing font (family, color, size), background (color, opacity), window (color, opacity), character (edge, style) and text opacity. Help support controls are provided as help, discussion forum, mail etc.

Video lecture Content is presented in different presentation styles. The activities are displayed in the progress bar. For example, color variation is used to differentiate, in-video activity inclusion, visited, playing video etc. Embedded quiz in the video are incorporated at different places either in between or at the end of video. Quiz further has various parameters that need to be considered. The quizzes differ in types, options available, question types etc. For example, quiz may be graded or non-graded; different controls in quiz; type of questions in quiz - MCQ, descriptive and true/false; correct answer response for the attempted question, and many more. The presentation style of the video interface and video lecture content are summarized in Table 3.

Table 2. The weighted feature checklist of video lecture of some MOOC providers.

	Feature	Control			Coursera	edX	Udacity	Weight
Video Interface	Display Time	Cuirrent Time			✓	✓	✓	3
		Total time			✓	✓	✓	3
	View Setting	Play/Pause			✓	✓	✓	3
		Volume Slider			✓	✓	✓	3
		Full screem mode			✓	✓	✓	3
		Modify Speed			✓	✓	✓	3
		Caption On/Off			✓	✓	✓	3
		Navigation			✓	✓	✓	3
	Advance Setting	HD support			X	✓	✓	2
		Poster Frame			X	X	✓	1
		Change video player			✓	X	X	1
		Caption multilingual support			✓	X	✓	2
		Change Caption format			X	X	✓	1
	Help Support	Report problem			✓	✓	✓	3
		Go to Discussion Forum			✓	✓	✓	3
		Keyboard Shortcut Help			✓	X	X	1
Video Lecture Content	Lecture Material	Download	Transcript	Srt	✓	✓	✓	3
				Pdf	X	X	✓	1
				Text	✓	✓	X	2
			Slides	Ppt	✓	X	X	1
				pdf	✓	✓	X	2
			Video (mp3/mp4)		✓	✓	✓	3
			Video quality		X	✓	✓	2
		Progress bar			X	X	✓	1
	Embedded Quiz	Availability			✓	X	✓	1
		Grading			✓	–	✓	2
		Control	Submit		✓	–	✓	2
			Skip		✓	–	X	1
			Continue		✓	–	✓	2
			Re-watch Instr.		X	–	✓	1
		Type	MCQ		✓	–	X	1
			Descriptive		X	–	✓	1
		Display Response			✓	–	✓	2
		Allowed attempts >1			✓	–	X	1
		Show Quiz presence			✓	–	X	1

Table 3. Presentation style of video lecture of some MOOC providers.

	Feature	Option	Coursera	edX	Udacity
Video Interface	View Setting	Speed	.75x-2.0x	.50x-2.0x	.25x-2.0x
		Navigation	Prev./Next	List	Prev./Next, Prog. Bar
	Advance Setting	Player Option	Flash, Html5	–	–
		Caption Lang. (#)	9	1	More than 60
		Caption Format	–	–	Font, Window, Background
	Help Support	Report problem	Help	Discussion Forum	Discussion Forum
		Discussion Link on	Video Interface	Lecture Page	Lecture page
	Lecture Material	Progress	–	–	Progress Bar
Video Lecture Content	Quiz	Display Response	Cor./Incor	–	No
		Correct Answer	Attempt> max	–	In Transcript
		Attempt Allowed	3	–	1
		Show Quiz presence	Color	–	No
		Location	Anywhere	–	At end

8 Case Study

The features of the MOOC video has been applied to three open source MOOC platforms. *Sakai*[9], *Open edX*[10], *CourseBuilder*[11] are MOOC platforms of Sakai foundation, edX and Google respectively, whose video interface have been chosen for our study. Study is focused on the two components of video lecture named, Video Interface and Video Lecture Content, and features provided to a MOOC student.

Sakai is an educational software platform developed for higher education by University of Michigan. Since its release in 2005, being used by more than 350 world's great colleges and universities organizations of diverse profiles list; over 4 million students worldwide [19]. It uses video lecture mechanism to offer lecture content.

Open edX is an open source release of edX platform in 2013. It is founded by Harvard university and Massachusetts Institute of Technology (MIT) [20]. Universities and educational providers are using it freely to offer their own MOOCs. Many

[9] https://sakaiproject.org
[10] http://code.edx.org/
[11] https://code.google.com/p/course-builder/

websites and MOOCs are launched on Open edX, listed at [21]. For our study we analyzed the Stanford OpenEdX [22] which is running on Open edX platform.

CourseBuilder is a software used to offer online courses. It provides an opportunity to universities and educational to offer their own MOOCs. It runs on google infrastructure[23] and is used world wide for offering MOOCs. We studied University of Auckland, New Zealand [24] that is running on CourseBuilder platform. Evaluation has been performed by student and guest account on their demo site.

Table 4. The feature checklist of video lecture for open source MOOC platforms.

	Feature Level	Control			Sakai	Stanford OpenEdX	Univ. of Auckland CourseBuilder
Video Interface	Basic	Cuirrent Time			✓	✓	✓
		Total time			X	✓	✓
		Play/Pause			✓	✓	✓
		Volume Slider			✓	✓	✓
		Full screem mode			✓	✓	✓
		Modify Speed			X	✓	X
		Caption On/Off			X	✓	✓
		Navigation			X	✓	X
		Report problem			✓	✓	✓
		Go to Discussion Forum			✓	✓	✓
	Advance	HD support			X	✓	X
		Caption multilingual support			–	X	X
	Optional	Poster Frame			X	✓	X
		Change video player			X	X	X
		Change Caption format			–	X	✓
		Keyboard Shortcut Help			X	X	X
Video Lecture Content	Basic	Download	Transcript Srt		–	X	X
			Video mp3/mp4		✓	✓	✓
		Progress bar			X	X	X
	Advance	Download	Transcript Text		–	✓	X
			Slides Pdf		✓	X	✓
			Video Quality		X	X	X
	Optional	Download	Transcript pdf		–	X	X
			Slides ppt		✓	X	✓
		Progress bar			X	X	X
		Embedded Quiz	L2	Grading	–	–	✓
				Controls	–	–	X
				Display reponse	–	–	✓
			L1	Allowed attempts >1	–	–	X
				Show Quiz presence	–	–	X

Table 5. The presentation style of Video lecture in some open source MOOC platforms.

	Feature	Control	Sakai	Stanford OpenEdX	Univ. of Auckland CourseBuilder
Video Interface	View Setting	Speed	–	.50x-2.0x	–
		Navigation	Prev./Next	Using List	–
	Advance Setting	Player option	–	–	–
		Caption Lang. (#)	–	1	1
		Caption Format	–	–	–
	Help Support	Report problem	Mail	Problem	Mailing list
		Discussion Link on	Discussion Forum	Lecture page	Mailing list
Video Lecture Content	Lecture	Progress	–	–	–
	Quiz	Display Response	–	–	Cort./Incor
		Correct Answer	–	–	Instant/End
		Attempts Allowed	–	–	1
		Show Quiz presence	–	–	No
		Location	–	–	Anywhere

Our weighted feature checklist is applied to video lecture components for different levels features provided by some open source MOOC platforms. In our case study, we applied our analysis for *Video Interface* and *Video Lecture Content,* and their *presentation style* on three MOOC platforms.

Table 4 and 5 displays a comparative feature checklist for each video lecture component and their presentation styles.

Some of the key findings are discussed here. *Display time*, one of the very common options is not provided by Sakai. *Caption* is also not present in *Sakai. Stanford OpenEdx* and *University of Auckland CourseBuilder* provide *advance settings* to better control of the video lecture interface while *Sakai* do not provide this facility. All platforms allow help support control for reporting the problem but in different ways. Downloading of the *downloadable transcript* is available only in *Stanford OpenEdX* but limited to text format; While *University of Auckland CourseBuilder* allows only viewing the *transcript. Different quality* of the video and *progress bar* is not supported by any provider studied. CourseBuilder is the only one providing embedded quiz.

Table 6 shows University of Auckland CourseBuilder provide most of the features for MOOC videos to their students for video interface as well as content. Also, the presentation style used for video lecture content that is very limited for platforms other than the University of Auckland CourseBuilder. Table 6 and Table 7 list the percentage of features supported and their level respectively, by the open source platforms mentioned in our study..

Fig. 2 displays coverage of *Video Interface* and *Video Lecture Content* components of video lecture in MOOC in our case study. Some of the key findings are as follows:

- The maximum controls are provided by CourseBuilder.
- All view setting options are available in Open edX.
- Advance settings are not present in Sakai.

- Each MOOC provider uses a similar number of controls for help support.
- Embedded quiz are only present in CourseBuilder video lecture.

Fig. 3 shows the provided features by some the providers in different levels. Some of the key observations about these are:

- The maximum basic features are provided by Stanford OpenEdX.
- Stanford OpenEdX is the only one that provided advance features or video interface.
- Optional features for video interface are not available in Sakai.
- The basic and advance features of video lecture content are equally supported by all the platforms, in our case study.
- The number of optional features available in University of Auckland CourseBuilder is double, in comparison to the other providers in our study.

Table 6. Percentage of features supported by some open source MOOC platforms.

Component	Controls	Sakai	Stanford OpenEdX	Univ. of Auckland CourseBuilder
Video Interface	Display Time	50%	100%	100%
	View Setting	50%	100%	66%
	Advanced Setting	0%	40%	20%
	Help Support	66%	66%	66%
Video Lecture Content	Lecture	38%	25%	37%
	Embedded Quiz	0%	0%	50%

Table 7. Percentage of feature level by some open source MOOC platforms.

Component	Level of Feature	Sakai	Stanford OpenEdX	Univ. of Auckland CourseBuilder
Video Interface	Basic	60%	100%	80%
	Advance	0%	50%	0%
	Optional	0%	25%	25%
Video Lecture Content	Basic	50%	50%	50%
	Advance	33%	33%	33%
	Optional	12%	12%	24%

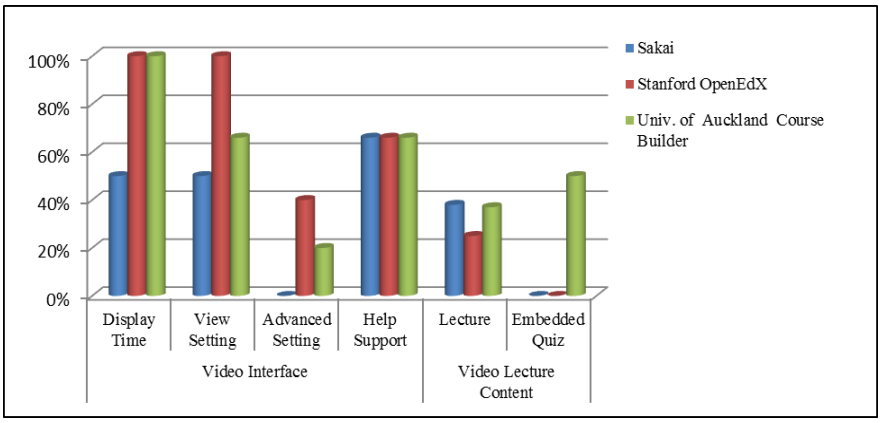


Fig. 2. Graph showing percentage features used by some open source MOOC platforms.

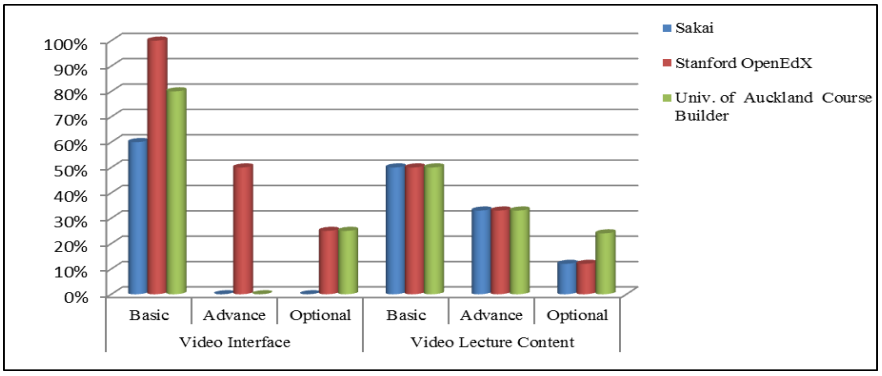


Fig. 3. Graph showing percentage of feature level used by some open source MOOC platforms.

9 Benefits

The features we presented provide aid for including videos in MOOC. For deriving our list, most popular MOOC platforms has been studied. Our list of features benefits the developers of MOOC for incorporating lecture video in their platform. We see that neither of the platforms provides all the features of video to their students.

Mostly, ad-hoc approach is used for providing video lecture to their students. Either the existing open source platforms are used which have their own mechanism for video lecture that needs to be extended sometimes or developing a new platform that needs identification of features that need to incorporated in video lecture. In both the cases, there is high probability of generating basic requirements of features. Developer of MOOC may not be able to elicit the options for video lecture.

The feature checklist presented here can be used by MOOC developers. It lists the features available in MOOC videos that need to be available for the student. During the feature elicitation, the developer can use the checklist to choose the desired options, with less effort. It facilitates the developer to include more options for videos. The video lecture mechanism developed using our checklist provides more interactivity to the interface and make it student oriented.

10 Limitations

For our analysis, we studied the most popular MOOC platforms. There may be some options that are provided by other MOOC providers that have not been studied. However the new options can be easily included in the checklist. Also the derived analysis is for video in MOOC. It is insufficient for standalone video interface creation. It may require some for options and features for standalone video interfaces, which are out of scope of this paper.

11 Conclusion

In this paper, we present a weighted checklist of features for the videos in MOOC. The checklist covers the features including different presentation styles of components of video lecture that are video interface and video lecture content. The requirement checklist is useful during incorporating videos features in MOOC. It eases the task of requirement specification for video software in MOOC by selecting and choosing the desired requirements of features from the checklist. The checklist presented here is extensible in nature and can be updated easily to add any new feature and option.

References

1. Voss, B.D.: Massive Open Online Courses (MOOCs): A Primer for University and College, Board Members, http://agb.org/sites/agb.org/files/report_2013_MOOCs.pdf (2013)
2. Wikipedia, http://en.wikipedia.org/wiki/Massive_open_online_course (2014)
3. Alario-Hoyos, C., Sanagustin, M.P., Kloss, C.D., Rojas, I.G., Leony, D.: Designing Your First MOOC from Scratch: Recommendations After Teaching "Digital Education of the Future", www.openeducationeuropa.eu/en/elearning_papers (2014)
4. Breslow, L.B, Pritchard, D.E., DeBoer, J., Stump, G.S., Ho, A.D., Seaton, D.T.: Studying learning in the worldwide classroom: Research into edX's first MOOC. In: Research & Practice in Assessment 8(1), pp. 13-25 (2013)
5. Seaton, D.T., Bergner, Y., Chuang, I., Mitros, P., Pritchard, D.E.: Who does what in a massive open online course?. 57, pp. 58-65 (2013)
6. Shah, D.: Class central report, https://www.class-central.com/report/coursera-10-million-students/ (2014)
7. Smith, L.: EducationDIVE, 5 education providers offering MOOCs now or in the future, http://www.educationdive.com/news/5-mooc-providers/44506/ (2012)

8. The New York Times, Article: The Year of the MOOC, http://www.nytimes.com/2012/11/04/education/edlife/massive-open-online-courses-are-multiplying-at-a-rapid-pace.html?pagewanted=all&_r=0 (2012)
9. The New York Times, Article: the Big three at a glance, http://www.nytimes.com/2012/11/04/education/edlife/the-big-three-mooc-providers.html?_r=0 (2012)
10. Kizilcec, R.F., Piech, C., Schneider, E.: Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In: Third International Conference on Learning Analytics and Knowledge, pp. 170–179. ACM, New York, USA (2013)
11. Sinha, T., Jermann, P., Li, N., Dillenbourg, P.: Your click decides your fate: Inferring Information Processing and Attrition Behavior from MOOC Video Clickstream Interactions. In: Proceedings of the 2014 Empirical Methods in Natural Language Processing Workshop on Modeling Large Scale Social Interaction in Massively Open Online Courses (October 2014)
12. Guo, P.J., Reinecke, K.: Demographic Differences in How Students Navigate Through MOOCs. In: First ACM conference on Learning@ scale conference, pp. 21-30. ACM, USA (2014)
13. Anderson, A., Huttenlocher, D., Kleinberg, J., Leskovec, J.: Engaging with Massive Online Courses. In: 23rd international conference on World wide web International World, pp. 687-698, Wide Web Conferences Steering Committee, Seoul, Korea (2014)
14. Guo, P. J., Kim, J., Rubin, R.: How Video Production Affects Student Engagement: An empirical study of MOOC videos. In: First ACM conference on Learning@ scale conference, pp. 41-50. ACM, USA (2014)
15. Chorianopoulos, K., Giannakos, M.N.: Usability design for video lectures. In: 11th European conference on Interactive tv and video, pp. 163-164. ACM Press, New York, USA (2013)
16. Kim, J., Guo, P.J., Seaton, D. T., Mitros, P., Gajos, K.Z., Miller, R.C.: Understanding In-Video Dropouts and Interaction Peaks in Online Lecture Videos. In: first ACM conference on Learning@ scale conference, pp. 31-40. ACM, USA (2014)
17. Ortega, Sergio, Francis, Brouns, F., Gutiérrez, A.F., Fano, S., Tomasini, A., Silva, A., Rocio, V. et al.: D2. 1 Analysis of existing MOOC platforms and services (2014)
18. Clark, R.C., Mayer, R.E.: E-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning. John Wiley & Sons, San Francisco (2011)
19. Sakai overview, https://sakaiproject.org/overview (2014)
20. Open edX, http://code.edx.org/
21. GitHub, https://github.com/edx/edx-platform/wiki/Sites-powered-by-Open-edX
22. Stanford online, http://online.stanford.edu/openedx
23. Course-builder checklist, https://code.google.com/p/course-builder/wiki/CourseBuilderChecklist
24. The University of Auckland[NZ], https://www.coursebuilder.cad.auckland.ac.nz

Using Fuzzy Logic in Knowledge Tests

Aleksandr Alekseev [1, *], Marika Aleksieieva [2], Kateryna Lozova [1], Tetiana Nahorna [1]

[1] Sumy State University, Faculty of Technical Systems and Energy Efficient Technologies, Sumy, Ukraine
[2] Harvard University, Graduate School of Arts and Sciences, Cambridge, Massachusetts, USA
alekseev_an@ukr.net, maleksieieva@fas.harvard.edu, katarina_lozovaya@ex.ua, nagorna-t@mail.ru

Abstract. An article describes the specialties of nonlinear scale formation of coincidence of standard answer with student's answer, basing on application of fuzzy logic during the test control of knowledge. It is given the detailed exposition of the mathematical apparatus that is used for substantiation the decision-making on the formation of the coincidence scale of answers. The author notes that using of the coincidence scale of answers gives the student an opportunity to express doubt and specify any degree of true answer ranging from "False" to "True". In this case test results are measured in the opposite terms from clear to fuzzy logic when the final mark is determined by the match of the answers. For example, if reference answer is equal to student's one it means that he/she knows the materials, and vice versa if the reference answer does not match, student does not know the topic. There are some types of the test tasks in the testing with using the coincidence scale of answers. The article describes the peculiarities off the parameter assignment of strictness the fuzzy-logic system. The results of experimental verification of the proposed innovations' effectiveness are given. These results allow stating the improvement of the measurement capabilities off the test with using the coincidence scale of answers basing on application of fuzzy logical calculations.

Keywords: Pedagogical measures, Knowledge test control, Measuring scale, Fuzzy logic, Strictness parameter, Test questions.

Key Terms: ICT Tool, Quality Assurance Process, Teaching Methodology, Teaching Process, Technology

1 Introduction

Fast and accurate evaluation of knowledge formation remains is a relevant task for long-existing forms of learning. Moreover, it has become an increasingly important for the recently emerged distance learning or blended learning (partial implementation

[*] Professor of Department of Manufacturing Engineering, Machines and Tools, Doctor of Pedagogy., Associate Professor

of distance learning technologies into classes that are conducted traditionally). The most important characteristics of the different forms of learning remains objective monitoring of students' academic achievements and construction of effective teaching methods based on it.

Further development of the theory and practice of the test control gives significant prospects for achievement of such goals. Using the information and telecommunication technologies, the test control successfully completes and improves existing traditional forms and methods of knowledge control. Computerized testing carries out a number of pedagogical functions assigned to knowledge test control, and becomes an effective means of summarizing the results of learning at all stages of education, from an entrance test to a comprehensive final exam.

2 Preconditions for using the fuzzy logic in the scaling of students' answers

Educational measurement technologies are developing in the close cooperation with the achievements of pedagogy, psychology, sociology and other empirical sciences, which are characterized by using the quantitative and quality indicators, differ by levels of manifestation of properties that are not measured directly. Due to this, there is no exception in the development of scaling tools applied to interpret the student's responses in higher education institutions with a computerized test control of knowledge.

The problem connected to the need in making available for the respondents involved in the questionnaire, or student who participates in the test control of knowledge, scale transfer of their judgments about the object of evaluation in the quantitative description of the level of assessed property has been known for a long time. Currently, there are several solutions proposed by different authors.

One of the first solutions was proposed by L. Thurstone [9]. The procedure of constructing the L. Thurstone scale is to provide an opinion about the level of assessed property in the frame of a set of evenly distributed judgments. Text description of each judgment is assigned a value of the level bar graph properties which corresponds to an interval scale. The scale constructed in such way is an interval one and its usage gives the possibility to apply a sufficiently wide range of statistical methods of processing the measurements results. However, a large amount of preparatory work related to the construction of the interval scale, relative equality of intervals, limit the possibilities of its application for the evaluation of students' knowledge.

The scale, developed by R. Likert [5], suggests the existence of the alternative judgments that reflect extreme levels of the assessed properties. These judgments can be formulated as "strongly agree" through "uncertainty" to "strongly disagree" for the test control. In addition, R. Likert scale bar graph set intermediate values associated with the specific levels of the assessed properties. In the text description no more than three of these intermediate values are commonly used, for example, "Somewhat agree", "neither yes nor no," "Somewhat agree".

The R. Likert scale is an ordinal scale, and despite the fact that, usually, its construction does not require the time-consuming preliminary work in the practice of educational measurement finds limited application. This is due to the fact that within the ordinal scales we can only arrange objects in ascending or descending order estimates of measurable properties at the lowest possible statistical treatment of the results of evaluations. Besides, the accuracy of pedagogical measurements with the use of scales that were developed due to approaches of R. Likert including numerical grading based scales limited by number of intermediate values, the number of which usually does not exceed 10-15.

Despite significant progress, reached for the last years in the different field of knowledge in the developing sphere, systematization and the field of analysis the results of practical application the methods of scaling the properties of qualitative and quantitative indicators, we have to realize that new approaches have limited application for the interpretation the student's responses on the knowledge test control.

To improve the accuracy of the pedagogical measurement, including the empowerment of statistical processing of the measurement results, it is necessary, in our view, to use such benefits in scaling the student responses using the ratio scale (name, by the definition of S.S. Stevens [8]). The mathematical apparatus of fuzzy logic will help to do things mentioned above.

3 Basic Provisions

During a traditionally organized examination the roles of a teacher and a student are allocated in accordance with the objectives of the oral control, when the teacher asks questions in order to identify student's generated knowledge. The student comprehends the questions and gives the answers, based on his/her idea of correctness of an answer. Then the teacher makes a judgment based on the results of such statements about the success of student's answers to particular questions, also evaluates the knowledge of all studied materials considering the total number of answers. Herewith, evaluating student's knowledge the teacher generally takes into consideration not only the formal correctness of the answers, but also how they were given, and whether the student was sure about the answers vs. showed a sign of insecurity, which may indicate instable knowledge. At the same time the student may use interpersonal contacts and consciously or unconsciously formulate the answer in such a way that it would let the teacher trace the causes of the seemingly unsuccessful answer. The doubt expressed in the answer provides an experienced teacher with another information channel that allows proper evaluating of the actual level of student's knowledge.

During classically organized test control, unlike oral examination, alienation of teacher's individuality occurs. Due to this fact, it is impossible to apply diagnostic capabilities of the teacher during the control process in order to identify the actual knowledge of the students.

Test control usually requires performing a task by selecting one of the possible answers or giving an unequivocal answer formulated from a limited set of words, letters, numbers, or graphics. In any case, the student should use his/her own experience and

make such an answer, which would contain the conclusion of true judgment in terms of strict logic. However, it is not possible to express doubt or indicate how the answer may differ from the correct one.

Checking results of the written test control the teacher has only a report of a student containing no data about possible difficulties in formulating the response. Therefore, the final evaluation cannot indicate whether the student was sure in the answer, or just speculated it relying only on luck. Computerized test control is more formal and matches the reference answers with the student's ones.

Concerning that, the developed test control simulation model [1] proposes to perform computerized control of knowledge by using of an expert system (Fig. 1) based on fuzzy logic [4]. Application of this system gives to a student the opportunity to operate not only the classical values of logical variables like "false" and "true", but also to use their intermediate values fading from one extreme value ("false") to an opposite one ("true").

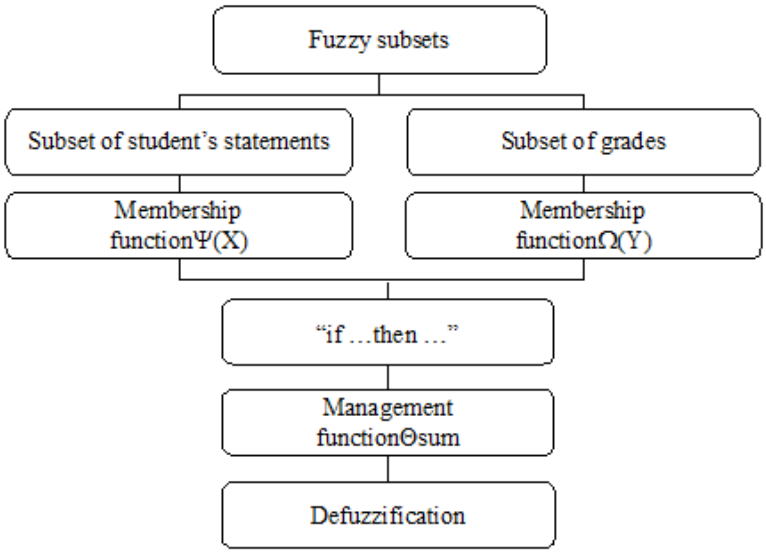


Fig. 1. Fuzzy logic expert system

The expert system uses piecewise continuous membership functions in order to define how evaluation of student's knowledge and his/her expressed statement relate to fuzzy logic subsets. These functions have transitional areas presented as segments a-b and b-c, connecting zero and one (maximum) levels of reliability (Fig. 2).

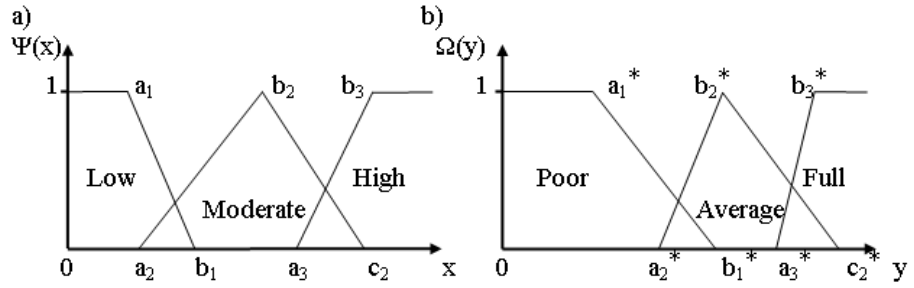


Fig. 2. Membership functions for subsets of student's statements (a) and evaluation of academic achievements (b)

A membership function for each term of the base term set of a logical variable "Level of matching answers" (Ψ(x)) is shown in Fig. 2,a. According to the mentioned above chart, all possible values of the function Ψ(x) are characterized as low, moderate or high level of matching answers depending on how the student's answer is close to the reference one. At the same time, mismatch of the answers can be caused not only by the incomplete knowledge, but also by insufficient confidence in knowledge, excessive emotions, or any other reasons preventing the student from making an unequivocal judgment about trueness of his/her conclusions.

Similar situation is with a membership function of a logical variable "Evaluation of student's knowledge" (Ω(y)), where the terms of a base term set are characterized by three gradations –"poor", "average" and "full" (Fig. 2,b) – depending on how the evaluation of the answer is close to one of evaluation scale criteria.

The presence of unrelated fuzzy logic sets allows to make such relevant fuzzy statements as "if ... then...". For example, clear logic accepts only two extreme statements: "if student's answer does not match the reference answer, then student's knowledge is unsatisfactory" and "if student's answer matches the reference answer, then the student has necessary knowledge". Fuzzy logic accepts both these extreme values, as well as any other intermediate statement linking the certain degree of answer accuracy and the corresponding answer evaluation.

The matching of subset items of the postulating and stating parts of a statement may apply a control function that is based on either "correlation –product encoding" method or "correlation – min encoding" method [4]. Currently there are no evidences confirming the preference of using one of these methods in computerized control of knowledge. However, "correlation – product encoding" method is used in the simulation model due to a number of reasons.

Sum combination method is used for getting a generalized logical statement. Herewith, superposition of membership functions of fuzzy sets is defined as

$$\Theta_{sum}(Z)=\Theta_i(Z) \quad \forall Z, \quad i\epsilon[1,3] \tag{1}$$

Transformation of a fuzzy set into a single decision taken on the basis of fuzzy logic statements requires using the gravity center of the fuzzy set membership function – centroid defuzzification method.

4 Strictness Parameter

The application of fuzzy logic relieves the student from necessity to speculate if he/she is not sure in the answer. Clearly indicating the degree of trueness in the answer, the student thereby provides the data giving possibility of mathematically differentiation of his/her academic achievements with high accuracy, and to perform unambiguous evaluation.

Mathematical application of fuzzy logic to the test control can also enter a "strictness" parameter. At the oral examination the teacher can somehow "forgive" a controversial answer deviating from his/her idea of trueness. But a stricter teacher will punish this controversial answer by a worse grade. Similarly to a traditional examinations conducted by teachers with different ideas of perfect knowledge of materials, the tests based on fuzzy logic may also be evaluated in different ways.

Fig. 3 shows an example of control which lets the student to give answers in relatively simple way in terms of fuzzy logic, if it is added to the test software interface. Indicating the degree of student's answer deviation from the reference one requires moving a slider to any position between the leftmost ("False") and the rightmost ("True), and clicking "OK". The slider location is determined and converted into relative coordinates, which are used for further calculations in a fuzzy logic expert system.

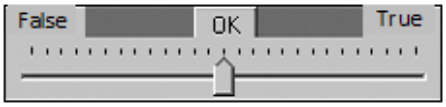


Fig. 3. Control of a fuzzy logic system

Rating answers and number of points accrued will depend on how the student indicated the degree of his/her answer matching the reference one. The number of points accrued depends also on the "strictness" parameter, which is indicated by sections in transition areas of membership functions –student's answer matches / does not match the reference answer, and the student learned / did not learn the controlled material. Despite the fact that the coordinating of these segments have quantitative indication, the level of strictness to student's knowledge is measured qualitatively, in such terms as "strict" (S), "stricter" (SS), "less strict" (LS), and "not strict" (NS), filling these concepts with quantitative measurement each time. Thus, if there is a necessity to compare the results of control, then introduction of the "strictness" parameter requires specified adopted coordinate values for transition sections.

Fig. 4 shows the charts illustrating changes in application rate of control for entering answers when implementing different "strictness" strategies of the fuzzy logic

expert system. In the diagrams was considered such data – we applied the information about the students who used the element of fuzzy-logic system in the computerized tests.

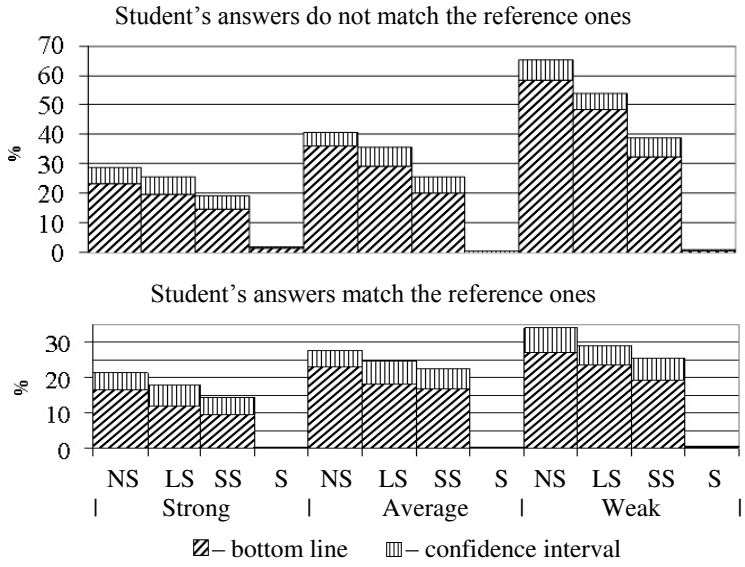


Fig. 4. Application rate of the control

The chart in Fig. 4 shows that in case of specified extreme level "Strict" the vast majority of students (over 98%), regardless of degree the preparedness (relatively strong, average and weak students) and degree of their answers matching with the reference ones, rarely use the opportunity to make a statement in terms of fuzzy logic. Absence of effective incentives upon the almost confident answer, and extremely punishable little doubt lead to the fact that students prefer to answer in terms of "False"-"True"." Thus, the fuzzy logic expert system capacities are practically not used. Therefore, it is not recommended to use a "strict" test system for practical purposes.

Other manifestations of "strictness" are popular enough to use fuzzy logic. It should be noted that the level "Not strict" is often demanded by weak students, as in case of student's answers matching the reference answers (30,5%±3,4%), and to even greater extent in case of answer mismatch (62,5%±4,1%). Therefore, this approach is not recommended to be a priority in order to ensure that all students are in equal conditions and none of them has any preference.

Table 1 shows the coordinates of the membership functions corresponding to the level of "Stricter". According to the chart in Fig. 4, it is often demanded and can be recommended as the core level in the absence of any other preferences. This recommendation can be confirmed by positive experience of use as the sole strictness parameter in the fuzzy logic system of test software SSUquestionnaire [7].

Table 1. Coordinates of membership function transition lines

	Ψ(X)							Ω(X)						
	Low		Moderate			High		Poor		Average			Full	
	a_1	b_1	a_2	b_2	c_2	a_3	b_3	a_1*	b_1*	a_2*	b_2*	c_2*	a_3*	b_3*
Y	0	1	0	1	0	1	0	0	1	0	1	0	1	0
X	0	0,25	0,15	0,5	0,85	0,75	1	0	25	15	50	85	75	100

Discussing the data shown in Fig. 4, it is necessary to underline that they do not directly recommend any of strictness degrees in the test system. So there may be different approaches to setting the "strictness" parameter. However, it is necessary to mention that regardless of the adopted approaches to setting the expert system strictness level, it must be set up prior to the test control. Any changes in the conditions of control through adjusting the strictness parameter for specific students, groups of students or disciplines are unacceptable. Like the oral examination, on the one hand there is contradiction between the desire to set up individual approach to each student and evaluate his/her achievements with the strictness degree that would enhance learning, vs. on the other hand, the requirement of compliance with the general approach to all students. Therefore, differentiating the "strictness" parameter in a fuzzy- logic test system can be justified for some special cases, but the general approach requires this parameter to be standardized, and academic achievement of any student should be equally evaluated, regardless of any subjective or objective circumstances.

5 Types of Tests

Despite the considerable variety of standardized test questions ([3], etc.), fuzzy logic expert system accepts only two types of tests.

The first type of tests covers the tasks containing the questions that can be answered using the full range of logical variables from "False" to "True". These are the questions that require to confirm or deny any statement, such as "Fish live in a water", "2 +2 = 4" or "The sun shines at night", "2 +2 = 5", etc.

When performing the test of the first type a student can move the control slider of a fuzzy logic expert system (Fig. 3) to any of the positions, which , in his/her opinion, corresponds to the degree of answer trueness. If one of the extreme positions is selected and student's answer matches the reference answer, the highest possible score will be awarded. If the selected extreme position of the slider does not match the reference answer, there will be 0 points. Intermediate position of the slider will allow giving intermediate (between zero and a maximum) number of points.

Another type of tasks includes the questions that can be answered within a half of the range of logical variables from "Not true" ("Not false") to "Truth". These tasks include questions along with two or more options of possible answers. At least one of them is correct and at least one is wrong. For example, if the task has a question "2 +2 =?" along with three answer options "3", "4" and "5", and it is offered to determine

which one is correct, then examinee cannot select the wrong answer “3” or “5” stating that it is false.

When performing such task, the control slider of the fuzzy logic expert system can be moved within a range from the middle position “Not true” (or “False”) to the rightmost position “True”. In this case, the maximum possible score will be given if student’s answer matches the reference one and the rightmost slider position is selected. In all other cases, the amount of points accrued will be determined by how student’s answer matches the reference one (depending on the slider position).

Table 2 shows different scoring options for the two considered types of tests (maximum score for correctly completed task is 100 points).

Table 2. Points accrued for a completed task depending on student’s answer matching the reference answer

Type of task	Student’s answer matches the reference one									
Type 1	False		¼		1/2		¾		True	
Type 2	½		5/8		3/4		7/8		1	
Strictness parameter value	Strict	Not strict	Strict	Not strict	Strict	Not strict	Strict	Not strict	Strict	Not strict
Answer evaluated, points	0	0	1	30	3	70	6	90	100	100

6 Measurement Capabilities

For the evaluation of the impact of a fuzzy logic expert system on the measurement capabilities of test knowledge control was made an experimental research.

The experiment engaged 228 students divided between the experimental and control groups. The groups were formed on the basis of current students’ progress. Mann-Whitney [6] checks showed that the groups are homogeneous.

The students in the experimental group were given a fully functional test program SSUquestionnaire, also they had an opportunity to give fuzzy logical answers. Strictness parameter of the fuzzy logic expert system was set up as “Stricter” and did not change throughout the experiment.

The test program used in the control group differed from the fully functional one, since its fuzzy logic module was disabled. The students could not move the control slider of the fuzzy logic expert system to any intermediate position; they had been forewarned as well.

Test results of the experimental and control groups were processed mathematically. They helped to estimate the strength of links between successful execution of individual test items and the final estimates the students received for all of the test questions. Pearson correlation coefficient [2] was calculated for the test results of each group independently. It was believed that the closer the absolute value of Pearson correla-

tion coefficient is to one, the tighter are links and measurement capabilities of the relevant test.

Comparison of the received data showed that the experimental group revealed closer linear dependence between the results of individual tasks and the general test results than the control group. Pearson correlation coefficient in the experimental group increased from 0,52 to 0,65 compared to the control group, that indicates better measurement properties of the test.

7 Conclusion

Elimination the identity of the person from the process of control enables using the diagnostic capabilities of the examiner during the test. This disadvantage of test control can be mitigated by use of an expert system developed on the basis of mathematical fuzzy logic.

The advantage of the fuzzy logic expert system hides in the fact that its introduction into a test program provides students with the opportunity not only to give the answers based on strict logic, but also to indicate any degree of answer trueness ranging from “False” to “True”. A student does not have to give a definite answer, even if he/she is required to go beyond the scope of their own knowledge. He/she can express doubt indicating how an idea of the true answer matches or does not match the reference answer. In this case, the test results are not measured in terms of clear logic (if the reference answer matches student’s answer, then the student knows the material, and vice versa), but in terms of fuzzy logic, when the final evaluation is determined by how these answers match.

The proposed justification of the decisions made by the examiner on the basis of the fuzzy logic expert system mitigates disadvantages of computerized testing as a tool for educational measurements, but does not eliminate these disadvantages entirely. Further efforts in the improving the theory and methods of test control, including methods directed on the fundraising the computer equipment for modeling diagnostic functions of the teacher in the control process will enhance the reliability of results of the evaluation of student’s knowledge.

References

1. Alexeyev, A. N., Alexeyeva, G. V.: A simulation model of the test control of knowledge and skills. *Computer-oriented educational system*, Kyiv, NPU. M. Dragomanov, No 7 (14), 65 - 71. (in Ukrainian) (2009)
2. Glass, G. V., Stanley, J.C.: Statistical methods in education and psychology – Englewood Cliffs, N.J., Prentice-Hall (1970)
3. IMS Global Learning Consortium. Accessible at http://www.imsglobal.org/question
4. Korneev V.V., Gareev A.F., Vasyutin S.V., Reich V.V.: Databases. Intelligent Processing of Information. – Moscow: Knowledge (in Russian) (2000)
5. Likert R., Roslow S., Murphy G. A.: Simple and Reliable Method of Scoring the Thurstone Attitude Scales. Journal of Social Psychology, 1934. Vol. 5, 228 – 238 (1934)

6. Mann, H. B., Whitney, D. R.: On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, 18, 50–60 (1947)
7. New Opportunities of Knowledge Testing using software SSUquestionnaire Version 4.10 Accessible at: http://test.sumdu.edu.ua (in Russian)
8. Stevens S. S.: Experimental Psychology. Moscow: Foreign Literature, Vol. 1 (in Russian) (1960)
9. Thurstone L. L.: The Measurement of Social Attitudes. Journal of Abnormal and Social Psychology, Vol. 26, 249 – 269 (1931)

Knowledge-Based Approach to Effectiveness Estimation of Post Object-Oriented Technologies in Software Maintenance

Mykola Tkachuk[1], Konstiantyn Nagornyi[1], Rustam Gamzayev[1]

[1] National Technical University "Kharkiv Polytechnic Institute",
Frunze str., 21, 61002 Kharkiv, Ukraine
{tka@kpi.kharkov.ua , k.nagornyi@gmail.com , rustam.gamzayev@gmail.com}

Abstract. A comprehensive approach to effectiveness's estimation of post object-oriented technologies (POOT) is proposed, which is based on structuring and analyzing of domain-specific knowledge about such interconnected and complex data resources within a software maintenance process as: 1) structural complexity of legacy software systems; 2) dynamic behavior of user's requirements; 3) architecture-centered implementation issues by usage of different POOT. The final estimation values of POOT's effectiveness are defined using fuzzy logic method, which was tested successfully at the maintenance case-study of real-life software application.

Keywords: post object-oriented technology, effectiveness, crosscutting functionality, knowledge-based approach, fuzzy logic.

Key terms: Software Engineering Process, Knowledge Representation, Decision Support, Model, Metric

1 Introduction: Problem, Actuality and Research Objectives

The most part of modern software systems are developed and maintained using object-oriented programming (OOP) [1]. Well-known and important problem to support such applications are often modifications on many their subsystems and development of new components to implement additional business logic due to new user requirements. In order to emphasize this issue we propose to use in this paper the notion "legacy software system" (LSS), similarly to the terms in software reengineering domain (see, e.g. in [2]). Permanent changes in LSS lead to design instability which causes a so-called crosscutting concern problem [3,4]. The OOP actually does not solve this issue, and usage of OOP-tools increases the complexity of an output source code.

During ten last years some post object-oriented technologies (POOT) were elaborated and became intensive development, especially the most known POOT are: aspect-oriented software design (AOSD) [5], feature-oriented software design (FOSD) [6] and context-oriented software development (COSD) [7]. All these POOTs utilize the basic principals of OOP, but in the same time they have additional features, which allow solving the crosscutting problem electively. From the other hand the usage of any POOT for LSS maintenance and reengineering is related to additional time and other efforts in software development. That is why many researchers emphasize the actual need to elaborate appropriate approaches to complex estimation of POOT's effectiveness usage in real-life software projects. It is additionally to mention that within the context of this paper we are talking about the POOTs which are focused on programming techniques exactly, but not about such software management trends as Extreme Programming (XP), Rapid Application Development (RAD), Scrum and some others [8], which also can be characterized as "post object-oriented" approaches.

Taking into account the issues mentioned above, the main objective of the research presented in this paper is to propose the intelligent complex approach to effectiveness's estimation of using POOTs in software maintenance. The rest of this paper is organized in the following way: Section 2 analyses some critical issues in OOP and reflects the phenomena of crosscutting functionality in software maintenance. In Section 3 the existing POOT are analyzed and the results of their comparing are shown with respect to software maintenance problems. In Section 4 we present the knowledge-based approach for effectiveness's estimation of POOT, which is based on structuring and analyzing of domain-specific knowledge about interconnected and complex data resources within a software maintenance framework. Section 5 presents first implementation issues and the results of test-case for the proposed approach. In Section 6 the paper concludes with a short summary and with an outlook on the next steps to be done in the proposed R&D approach.

2 Some Critical Issues in Object-Oriented Programming and Crosscutting Functionality Phenomenon in Software Maintenance

To meet new requirements existing LSS have to be refined with new classes, which must implement their new functionality. Standard OOP toolkit "proposes" to support additional associations between already existent and new program objects, to modify inheritance tree for classes, to implement new or additional design patterns, e.g. the Gang-of-Four (GoF) patterns [9]. Because of permanent modifications on source code and doing software system re-design, developers face with "bottlenecks" of OOP: increase coupling among classes [10]; increase of depth of inheritance tree (DIT) for class hierarchies [11]; modification of design pattern instances [12,13]; emerging lack of modularity in functionality realization [14].

A number of studies investigate problems of OOP mentioned above, and theirs negative influence on LSS maintenance. High dynamic of requirement changes and these critical issues of OOP induce and propagate an additional development problem: this is a crosscutting concern's phenomena. Crosscutting concern (hereby referred as

"crosscutting functionality" - CF) is a concern emerges on user requirements level and often crosscuts on design level, this is a part of a business logic, which can not be localized in the separate module on source code view but stays separate on requirement view [15]. In literature exists a lot of researches related to CF's properties, multiple patterns of CF and it's interaction with the source code of non-crosscutting functionality, and it's further propagation in system's source code (see e.g. in [13 - 16]). There are some widespread examples of software system features which could be consider as CF: exception management, logging, transaction management, data validation [17]. Nevertheless our own experience in software development and LSS maintenance exposes that almost any system feature, emerged by requirements, on source code perspective could be transformed into CF.

CF has two main properties [18]: scattering and tangling. CF's source code *scatters* among classes (components) of non-crosscutting functions, this happens because of mismatch on end user requirement's level of abstraction, and final realization of this requirement as a feature on the source code level. CF's source code *tangles* (mixes up) with source code of the other functionality, no matter crosscutting or non-crosscutting. Moreover CF could be divided into several types [19]: homogeneous and heterogeneous. *Homogeneous* CF represents the same piece of source code which crosscuts multiple locations in multiple OOP-classes of a software system. *Heterogeneous* CF represents each time unique piece of source code which crosscuts multiple locations in multiple OOP-classes of a software system (see Fig.1).

```
public class Line {
private Point p1, p2;
 Point getP1(){ return
p1; }
 Point getP2() { return
p2; }
 void setP1(Point p1){
 this.p1 = p1;
Display.update(this);}}
public class Oval {
void setPosition(Point
p2){
  this.p2 = p2;
Display.update(this);
}
}
```

// Homogeneous CF

```
public class
CreditCardProcWithLogging{
Logger _logger;
public void debit(CreditCard
card, Money amount)throws
InvalidCardException,
NotEnoughAmountException,
CardExpiredException {
 _logger.log("Starting
debiting"
 + "Card: " + card
 + " Amount: " + amount);
// Debiting
 _logger.log("Debiting
finished"
 + "Card: " + card); }
```

// Heterogeneous CF

Fig. 1. Crosscutting functionality types

As a result, a presence of the CF in software system increases a complexity of the maintenance process [20]:

- CF complicates traceability of various software design artifacts, e.g requirements traceability [21];
- CF decreases understandability of a source code and functionality it realizes;
- source code of LSS becomes redundant;

- Almost impossible to reuse CF solutions, because of lack of modularity.

A conceptual approach, which allows to deal with CF, is a separation of concerns (SoC) [22]. It envisages a *decomposition* and further non-invasive *composition* of CF source code with the rest code of LSS. Decomposition mechanism allows to split source code into fragments and to organize them into easy-to-handle CF-modules. Composition mechanism supports reassembling of isolated code fragments in easy and useful way. Usage of SoC principles makes possible to decrease coupling in LSS, to decrease code redundancy, to reuse isolated CF-modules, to configure system by add/remove functionality if needed.

Finally, the existing POOTs provide SoC principles and offer a lot of toolkits to manage CF-problem in an effective way.

3 Post Object-Oriented Technologies: Main Features and Results of Comparative Analysis

As already mentioned above (see in Section 1) nowadays there are 3 main well-defined approaches in POOT-domain, namely: aspect-oriented software development (AOSD) [5], feature-oriented software development (FOSD) [6] and context-oriented software development technology (COSD) [7]. In order to reflect their essential features with respect to the problem of CF it is useful to represent an interaction between basic components of OOA and POOT [20].

AOSD was proposed in Research Center Xerox/PARC and it is now implemented in many programming languages such as Java / AspectJ, C ++, .NET, Python, JavaScript and some others [4]. AOSD allows to concentrate CF in separate modules called *aspects*, which should be localized in source code infected with CF using such means as points of *intersection* (point-cut) and *injection* (injection). Schematically this interaction is shown in Fig. 2, (a), where the white vertical rectangles C1, C2, C3 represent OOP-classes and gray horizontal rectangles A1, A2, A3 represent the aspects.

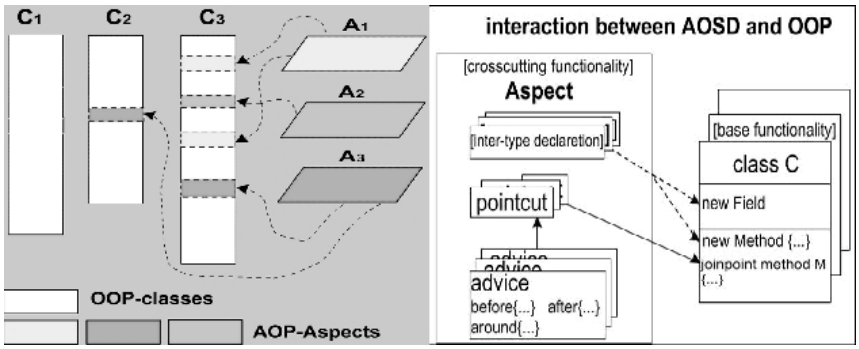


Fig. 2. AOSD: (a) – the conceptual scheme; (б) – the implementation facets (compare with [19])

More detailed the structure of aspect is shown in Fig. 2, (b). Any *aspect* consists of interconnected point-cut, of a *notification* (advice), and of an *introduction* (inner declaration). The task of *point-cut* is to define a connection point between *aspects* and basic methods in OOA-classes, in other words, *point-cut* determines those lines of code in the OOA-methods, were *notification* code has be introduced. A *notification* is a peace of code in OOA-language (e.g. in Java), which implements an appropriate CF, therefore notifications can be of three types: *before* – such a notification is perform before to call a OOA-method; *after* - a notification is made after this call; and *around* - a notification is executed instead to call a OOA-method. Also AOSD allows the introduction in OOA-classes new fields and methods that can be defined in aspects.

In the same way the FOSD and COSD schematically can be represented and analyzed carefully (see in [20] for more details). The results of this comparative analysis are presented in the Table 1.

Table 1. Results of comparative analysis for different POOT

POOT features / Estimation marks	Type of POOT		
	AOSD	FOSD	COSD
Modeling CF features at a higher level of abstraction	+	+	+
Implementation of homogeneous CF	+	+/–	+/–
Implementation of heterogeneous CF	+/–	+	+
Provide CF layers separately from a OOA-class	+	+	+
Context-dependent activation/deactivation of layers	–	–	+
Possibility to use several approaches simultaneously	+ /–	+ /–	–
Availability of CASE-tools to support this POOT	+	+	+ /–

Even a cursory analysis of this comparison shows that for a decision on the appropriateness and effectiveness of using an appropriate POOT to solve CF-problem in given LSS, it is necessary to take into account a number of other additional factors, which will be considered in the proposed approach.

4 Knowledge-Based Framework for Effectiveness's Estimation of Post Object-Oriented Technologies

Taking into account the results of performed analysis (see Section 2), and basing on some modern trends in the domain of POOT-development (see Section 3), we propose to elaborate a knowledge-based framework for comprehensive estimation of POOT-effectiveness to use them in software maintenance. Thus we proceed from one of possible definition of the term "knowledge" within the knowledge management domain [23], namely: *a knowledge is a collection of structured information objects and relationships combined with appropriate semantic rules for their processing in order to get new proven facts about a given problem domain.*

Then our next task is to define and to structure all information sources, and to elaborate appropriate algorithms and tools to process them with respect to the final

goal: how to estimate usage effectiveness of different POOTs in software maintenance.

4.1 Multi-dimensional model for POOT effectiveness's estimation

To implement the proposed knowledge-based approach the multi-dimensional modeling space is proposed in [20], and its graphical interpretation is shown in Fig. 3. According to this model the integrated effectiveness level is depend on two main interplaying factors, namely: 1) what type of LSS has to be modified with usage of an appropriate POOT; 2) what kind of POOT is used to eliminate the CF in this LSS. In order to answer these questions the following list of prioritized tasks can be composed:

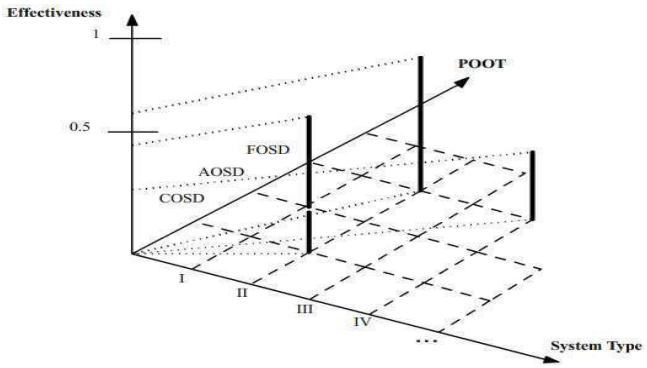


Fig. 3. 3-D modeling space for POOT's effectiveness estimation

(i) to define a type of given LSS with respect to its structure complexity and to behavior of requirements, which this LSS in maintenance process is facing with;
(ii) to calculate an average effort values for different POOT, if this one is used to eliminate CF in an appropriate LSS;
(iii) to elaborate the metrics for CF assessment before and after LSS modification using a given POOT;
(iv) to propose an approach to final effectiveness estimation of POOT's usage taking into account the results provided by activities (i) – (iii).

Below these tasks are solved sequentially, using knowledge-based and expert-centered methods and tools.

4.2 Definition of legacy software system types

To solve task (i) from the their list given in Section 4.1 the approach to analyzing and assessments of LSS's type proposed in [24] can be used, which is based on the following terms and definitions.

Def#1. *System Type* (ST) is an integrated characteristic of any LSS given as a tuple:

$$ST =< Structural\,Complexity, Requiremen\,tRank > \quad (1)$$

The first parameter estimates a complexity level of a given LSS, and the second one represents status of its requirements: their static features and dynamic behavior.

To calculate structural complexity (SC) the following collection of metrics was choused: *Cyclomatic Complexity* (V), *Weighted Method Complexity* (WMC), *Lack of Cohesion Methods* (LCOM), *Coupling Between Objects* (CBO), *Response For Class* (RFC), *Instability* (I), *Abstractness* (A), *Distance from main sequence* (D). The final value of SC can be calculated using formula (2), where the appropriate weighted coefficients for each metric were calculated in [24] with help of Analytic Hierarchy Process method [25].

$$\begin{aligned} SC &= K_V avgV + K_{WMC} avgWMC + K_{LCOM} avgLCOM + K_{CBO} avgCBO + \\ &+ K_{RFC} avgRFC + K_I avgI + K_A avgA + K_D avgD \end{aligned} \quad (2)$$

To evaluate the final value of SC of given LSS in terms of an appropriate linguistic variable (LV): *"Low", "Medium", "High",* the following scale was elaborated [24]:

$$\begin{aligned} SC_{Min} \le Low < \frac{2*SC_{Min} + SC_{Max}}{3} \\ \frac{2*SC_{Min} + SC_{Max}}{3} \le Medium \le \frac{SC_{Min} + 2*SC_{Max}}{3} \\ \frac{SC_{Min} + 2*SC_{Max}}{3} < High \le SC_{Max} \end{aligned} \quad (3)$$

To define the second parameter given in formula (1), two relevant features of any requirement were considered [24], namely: a grade of it's *Priority* and a level of it's *Complexity*.

Def#2. *Requirements Rank* is a qualitative characteristic of LSS defined as a tuple:

$$Requiremen\,tRank =< Priority,\ Complexty > \quad (4)$$

In [24] is mentioned that in modern requirement management systems (RMS) like IBM Rational Requisite Pro, CalibreRM and some others, the *Priority* and *Complexity* of requirements are usually characterized by experts in informal way, e.g. using such terms as: *"Low", "Medium", "High"*. The real example of such interface in RMS is presented in Fig. 4, with requirement's attributes "*Priority*" and "*Complexity*" (or "Difficulty" in terms of RMS-technology).

Taking into account the definition for linguistic variable (LV) given in [26], the appropriate term-sets for LVs *Priority* and *Complexity* respectively were defined in [24] as follows:

Requirements:	Priority	Difficulty	Stability
SR1: Parse Java Code	High	High	Medium
SR2: Elaborate Lexer for Java 5	High	Medium	Medium
SR3: Recognize all java lexical structures	High	High	Medium
SR4: Possibitity to parse single file	Medium	Low	Medium
SR5: Possibitity to parse whole package	Medium	Medium	Medium
SR6: Collect code statistics	Low	Medium	Medium
SR7: Recognize Java Grammatic	High	High	Medium
<Click here to create a requirement>	Medium	Medium	Medium

Fig. 4. The list of requirements completed in RMS Rational Requisite Pro

$$X : Priority \;;\; T(Priority) = \{"neutral","actual","immidiate"\} \quad (5)$$

$$X : Complexity \;;\; T(Complexity) = \{"low","medium","high"\} \quad (6)$$

Basing on definitions (1) – (6), the mapping procedure between 2 attribute spaces was elaborated in [24]. These attribute spaces are defined with apprpopriate LVs, namely: the space *"Requirements Rank"* with axes *"Priority"* and *" Complexity"*; the space "*System Type*" with axes "*Requirements Rank*" and "*Structural Complexity*". This mapping procedure in details is presented in [24], and the final result of this approach is shown on Fig. 5. It illustrates the main advantages of the proposed approach, namely: 1) we are able to estimate current state of system requirements w.r.t. their static and dynamic features; 2) basing on this estimation, we can define an appropriate type of investigated software system (e.g., some LSS in maintenance process), taking into account it's structural complexity and dynamic requirements behavior as well.

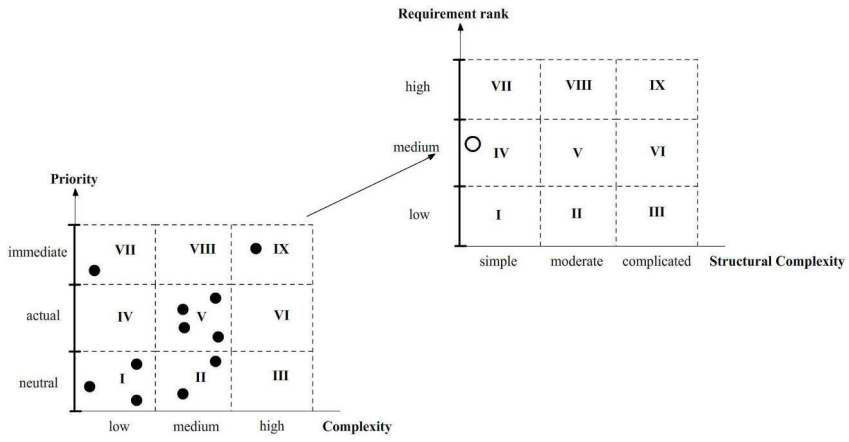


Fig. 5. (a) – the initial allocation of system's requirements in the space *"Requirement Rank"*; (b) – the mapped system's position in the space *"System Type"*

4.3 An architecture-centered method for POOT effort calculation

In order to solve task (ii) from their list given in Section 4.1 it is proposed to analyze basic architectural frames, which can be constructed for different POOT with usage of their OOP-specification. In [20] the following definition is proposed for this purpose.

Def#3. *Enhanced architectural primitive* (EAP) is a minimal-superfluous component-based scheme, which is needed to implement an interaction between basic OOP-elements (class, field, method) and specific functional POOT-elements.

Obviously, to perform the comparative analysis of different EAP in the correct way, they preliminary have to be represented in some uniform notation. As a such notation the architecture description language (ADL) should be used, because: 1) this notation is not depend on any specific programming tools; 2) in this way static and dynamic features of AP both can be described and analyzed.

The most important modeling abstracts of ADL (see e.g. in [27]) are *components*, *port*s and *connectors*, and there are such additional ADL - features as *role* and *interaction.* They have the following definitions within the context of this paper.

Def#4. *Component i*s a complex of functional items, which implements a certain part of a business logic in LSS, and which is supposed to have special interfaces (ports) for communication with other entities in an operating environment.

Def#5. *Port* is an interface to provide an interaction between several components.

Def#6. *Connector* is a special architectural item to join ports of different components.

Def#7. *Role* is a special feature of a given connector to identify its communicating ports.

Def#8. *Interaction* is a special feature of given connector defined using its roles.

More detailed the notion *port* can be characterized in the following way: 1) there is so-called *single port* - this is an interface of any component to communicate with some another one via exactly one connector; 2) furthermore there is a *case-port* - this is an interface of any component to communicate with another components via more then one connectors (e.g., using an appropriate Boolean variable as a flag to switch communication, etc.). Similarly, the notion *connector* can be classified as follows: 1) a *binary connector* – this is a connector with 2 fixed roles only; 2) a *multiply connector* – this is a connector, which has exactly 1 input role and more then 1 output roles; 3) a *case connector* – this kind of connectors can have a lot of input and output role as well.

Using the definitions Def#3 – Def#8 the appropriate EAP for all mentioned above POOT were elaborated [20]. As one example the EAP for AOP is shown on Fig. 6, which reflects how the specific AOP-features such as *advice* and *inner declaration* (they are shown as rectangular icons in grey color) are interacting with basic OOP – elements, namely: *class, field* and *method* (they are represented as crosswise icons in white color).

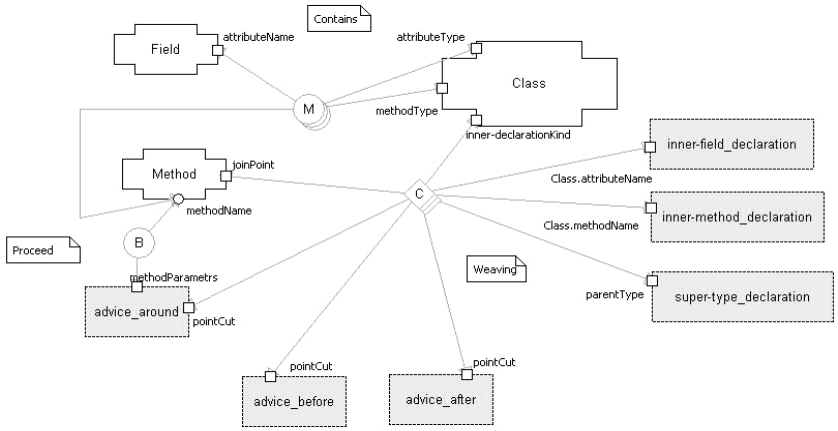


Fig. 6. ADL-specification for the aspect-oriented EAP

To calculate the complexity coefficients (CC) of the elaborated EAP the following formulas are proposed in [20], namely:

$$Component = 0.6*\#POOT + 0.4*\#OOP, \quad (7)$$

where a $Component$ is the CC of an appropriate EAP, $\#OOP$ is a number of architectural OOP – components, and $\#POOT$ is a number of POOT – components included in this EAP. These values are multiplied with the weight coefficients: *0,6* and *0,4* respectively, and these coefficients can be defined using some expert methods (see in [20] for more details);

$$Connector = 0.2*\#BinaryConnector + 0.3*\#MultyConnector + 0.5*\#CaseConnector, \quad (7)$$

where a $Connector$ is the CC of connectors included in an appropriate EAP, which is calculated using the number of binary connectors: $\#BinaryConnector$, the number of multi-connectors $\#MultyConnector$ and the number of case-connectors: $\#CaseConnector$, with respect to the appropriate weight coefficients *0.2, 0.3* and *0.5,* which also are defined by some experts [20];

$$Port = 0.8*\#SinglePort + 0.7*\#CasePort, \quad (8)$$

where $Port$ is the CC of ports included in an appropriate EAP, which takes into account the number of single ports: $\#SinglePort$, and the number of case ports: $\#CasePort$ with appropriate weigh coefficients.

Using formulas (7) – (9) the summarized value $Complexity$ of an appropriate EAP, measured in so-called architectural units (a.u.) [20] can be calculated as follows:

$$Complexity = Component + Connector + Port \quad (9)$$

The final values of CC for all POOT were calculated using formula (10), and they are represented in Table 2 (see in [20] for more details).

Table 2. The values of architectural complexity for the different POOT

POOT type	CC for components (a.u.)	CC for connectors (a.u.)	CC for ports (a.u.)	Summarized values of CC (a.u.)
AOSD	4,8	1	4,3	10,1
FOSD	3,6	1	3,9	8,5
COSD	2,8	0,7	4,1	7,6

Basing on the estimation values aggregated in Table 2 it is possible to make conclusions about average implementation efforts by usage of appropriate POOT to solve CF-problems in legacy software systems within their maintenance.

4.4 Quantitative metrics for crosscutting in legacy software

There are different ways to characterize a nature of the CF and it's impact to software source code. A number of studies are dedicated to a classification, qualitative and quantitative description of CF problem [3,14-16]. The aim of our research is to assess an impact, which CF makes to a structure of OOP-based software system during it's evolution in maintenance; therefore we are focusing on quantitative facet of crosscutting nature. To reach this goal it is proposed to perform next three steps.

Step 1: Localize source code belonged to a particular CF in a given LSS. Although exists several source code analysis tools for CF localization, e.g., tool CIDE [28], this problem remains really complicated for autoimmunization and demands an expert in code structure and business-logic of an appropriate LSS.

Step 2: Calculate a specific crosscutting weight ratio of a particular CF in the system indicated as CF_{ratio} [20]. This coefficient shows a ratio between OOP-classes, "damaged" by a particular CF and all OOP-classes in the system, or it's projection, e.g. business logic realization without subordinate classes of a framework. This coefficient possible to represent as

$$CF_{ratio} = \frac{C_{cf}}{C_{cf} + C}, \quad (10)$$

where C_{cf} – number of classes in LSS, "damaged" with CF, C – number of classes free of CF. Obviously, that $CF_{ratio} \in [0;1]$, and if $CF_{ratio} = 0$, it means a particular

functionality is not crosscutting; and if CF_{ratio} = 1, it means all classes are "damaged" with a particular CF.

Step 3: Calculate a residual crosscutting ratio indicated as RCR_{ratio}. This metric, based on DOS (Degree of Scattering) value, proposed in [14], namely "…*DOS is normalized to be between 0 (completely localized) and 1 (completely delocalized, uniformly distributed)*". Nevertheless this metric does not allow to asses "damage" degree, done by a particular CF, therefore we propose to refine DOS-metric in following way

$$RCR_{ratio} = DOS \cdot CF_{ratio}, \quad (11)$$

where DOS – Degree of Scattering; CF_{ratio} – specific crosscutting weight ratio of a particular CF. Similarly to CF_{ratio}, $RCR_{ratio} \in [0;1]$, if RCR_{ratio} = 0, it means that CF is localized in a separate module and it is no more crosscutting; if RCR_{ratio} = 1, it means that CF effects a whole system and is uniformly distributed.

Thus the proposed quantitative metrics (11) – (12) give to an expert a possibility to assess a distribution nature of a CF, and to estimate a "CF-damage" for a given LSS.

4.5 Fuzzy logic approach to complex effectiveness estimation of POOT

Based on assessment of POOT average implementation efforts (see Chapter 4.3), and assessment for residual crosscutting ratio (see Chapter 4.4) it is possible to estimate an integrated effectiveness of POOT usage. Although because of different scale and units of measurement for proposed assessments, it is hard to evaluate them within a single analytical method. Therefore, for further evaluations it is proposed to use one of algorithms of the fuzzy logic [26], namely the Mamdani's algorithm, which consists of 6 steps. According to this algorithm to estimate effectiveness of POOT usage it is necessary to compose fuzzy production rules (FPR). In this paper a verbal description for these rules is omitted, instead of this the widespread symbolic identifiers for short description of FPR are listed in Table 3.

Table 3. A symbolic representation form for the description for FPR

Symbolic form	Description
Z	Zero
PS	Positive Small
PM	Positive Middle
PB	Positive Big
PH	Positive Huge

The whole system of elaborated FPR consists of 20 definitions (see in [29] for more details), and the fragment of this FPR-system is listed below:

1. RULE_1: If "β_1 is **PS**" and "β_2 is **Z**", then "β_3 is **Z**";

2. RULE_2: If “ β_1 is **PM**” and “ β_2 is **Z**”, then “ β_3 is **Z**”;
3. …
4. RULE_9: If “ β_1 is **PS**” and “ β_2 is **PM**”, then “ β_3 is **PM**”;
5. …

Corresponding to the Mamdani’s algorithm, the next step is a fuzzifying of variables in FPR, therefore average implementation efforts, residual crosscutting ratio, and effectiveness of POOT usage have to be represented as LV. The output LV E_{POOT} is the effectiveness of POOT-usage, the LV E_{POOT} is bounded on universe X, and it belongs to the interval [0;1]. The term set for this LV looks like:

$E_{POOT} \in \{non-effective, low-effective, mid-effective, effective, very-effective\}$, and it could be represented in short form as $E_{POOT} \in \{Z, PS, PM, PB, PH\}$. The corresponding identifier for E_{POOT} is β_3 (see FPR above), and it is shown in Fig. 7.

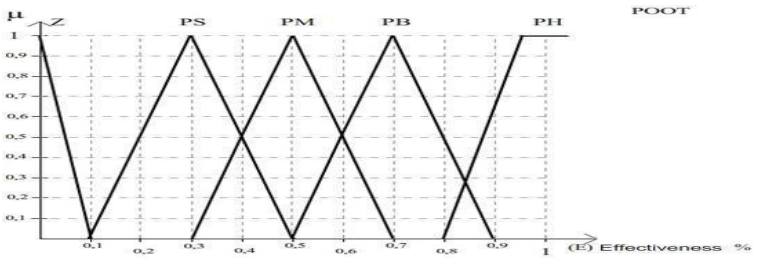


Fig. 7. The graphic form for LV “Effectiveness” E_{POOT}

The input LV C_{POOT} represents average implementation efforts, C_{POOT} is bounded on universe X and belongs to an interval [(EAP)min; (EAP)max], where EAPmin, EAPmax are minimum and maximum values of architectural complexity (measured in a.u.) for appropriate LSS type respectively. The term set for the C_{POOT} linguistic variable (LV) looks like: $C_{POOT} \in \{low, middle, high, huge\}$ and could be represented in short form $C_{POOT} \in \{PS, PM, PB, PH\}$. The corresponding identifier for C_{POOT} is β_1 (see FPR above). The graphical interpretation for this LV is similar to the graphic, depicted on Fig. 7.

The input LV P_{POOT} is a residual crosscutting ratio (see formula (12)). The LV P_{POOT} is bounded on universe X and belongs to interval [0;1]. The term set for this variable looks like: $P_{POOT} \in \{useless, low, middle, high, huge\}$, and it could be represented in short form as $P_{POOT} \in \{Z, PS, PM, PB, PH\}$. The corresponding identifier for P_{POOT} is β_2 (see FPR-system above). The visual interpretation is similar to the graphic depicted in Fig. 7.

5 The Test-Case and Result Discussion for the Proposed Approach

To illustrate the proposed approach the real LSS for personal data management was analyzed [29]. It consists of 15 java-classes, and it contains a homogenous realization of "logging" crosscutting functionality. Accordingly to the LSS – type definition method (see Section 4.2) this application belongs to the III-rd system type with rank: {"Low structural complexity"; "High requirement rank"}. The source code of this LSS was sequentially modified using 3 POOT: AOSD, FOSD, and COSD respectively. The final results of POOT effectiveness estimation are shown in Table 4. The first column lists all LSS – modifications to be compared: an initial OOP - version, which has to be re-structured wit respect to CF-problem, and its 3 modifications done with usage of different POOT. In the second column the summarized efforts needed for these modifications with respect to architectural-centered complexity are calculated (see Section 4.3). The data given in the third column of Table 4 show the level residual crosscutting ratio which is presented (for initial OOP-version) or which is remained after its redesigning with the appropriate POOT. The forth column indicates the final effectiveness's estimation values for all LSS-versions.

Table 4. Effectiveness of usage of POOT in a target system

(P)OOT	Architectural complexity (a.u.)	Residual crosscutting ratio (%)	Effectiveness level (%)
OOP	122.51	69.52	6,7
AOSD	79.43	0,15	73,3
FOSD	116.16	29.06	34,4
COSD	115.88	8.78	32,8

The results achieved show, that OOP actually is not enough effective to solve crosscutting problem (done with 6.7% only). The most preferable approach to eliminate this issue in the given type of LSS (as mentioned above, this is the III-rd system type according to LSS-classification proposed in Section 4.2), is an AOSD which provides effectiveness level over than 70%.

It is also to mention, although an effectiveness level of COSD and FOSD is lower than AOSD, over 30% for homogenous CF, it is still much better result than OOP. Taking into account a qualitative advantage of these two another technologies, namely: a possibility to implement a heterogeneous CF also (see Table 1), it can be reasonable to use one of them for LSS-maintenance to deal with such kind of CF in much effective way than AOSD.

6 Conclusions and Future Work

In this paper we have presented the intelligent approach to effectiveness's estimation of modern post object-oriented technologies (POOT) in software development, which

aims to utilize domain-specific knowledge for this purpose. This knowledge base includes such important and interconnected data resources as: 1) structural complexity of legacy software; 2) dynamic behavior of user's requirements; 3) architectural-centered implementation efforts of different POOT. To process these data the quantitative metrics and expert-oriented estimation algorithms were elaborated. The final complex estimation values of POOT's effectiveness assessment are defined using fuzzy logic method, which was successfully tested on some real-life legacy software applications.

In future we are going to extend a collection of metrics for POOT-features assessment, and to apply some alternative (to fuzzy logic method) approaches to final decision making. Besides that it is supposed to develop an appropriate software CASE-tool for expert's data handling in the proposed knowledge-based estimation framework.

7 References

1. Sommerville, I.: Software Engineering. Addison Wesley (2011)
2. Eilam, E.: Reversing: Secrets of Reverse Engineering. Wiley Publishing (2005)
3. Sven Apel et al. On the Structure of Crosscutting Concerns: Using Aspects of Collaboration? In: Workshop on Aspect-Oriented Product Line Engineering (2006)
4. Przybyłek, A.: Post Object-oriented Paradigms in Software Development: A Comparative Analysis. In: Proceedings of the International Multi-conference on Computer Science and Information Technology, pp. 1009-1020 (2007)
5. Official Web-site of Aspect-oriented Software Development community, http://aosd.net
6. Official Web-site of Feature-oriented Software Development community, http://fosd.de
7. Official Web-site of Context-oriented Software Development group, http://www.hpi.uni-potsdam.de/hirschfeld/cop/events
8. Highsmith, J.: Agile Project Management. Addison-Wesley (2004)
9. Gamma, E. et al. Design Patterns: Elements of Reusable Object-Oriented Software. Addison-Wesley (2001)
10. Sheldon, T., Jerath, Kh., Chung, H.: Metrics for Maintainability of Class Inheritance Hierarchies. J. of Software Maintenance and Evolution, Vol. 14, pp. 1--14 (2002)
11. Harrison, R. Counsell, S.J.: The Role of Inheritance in the Maintainability of Object-Oriented Systems. In: Proceedings of ESCOM '98, pp. 449--457 (1998)
12. Aversano, L. Cerulo, L. Penta, M. Di.: The Relationship between Design Patterns Defects and Crosscutting Concern Scattering Degree: An Empirical Study. J. IET Software, vol. 3, pp. 395--409 (2009)
13. Hannemann, J., Kiczales, G.: Design Pattern Implementation in Java and AspectJ. In: Proceedings of OOPSLA'02, pp. 161--173 (2002)
14. Eaddy, M. et al.: Do Crosscutting Concerns Cause Defects? In: IEEE Trans. Softw. Eng., 34(4), pp. 497--515 (2008)
15. Filman, R., Elrad, S. Aksit. M.: Aspect-Oriented Software Development. Addison Wesley Professional (2004)
16. Figueiredo, E.: Concern-Oriented Heuristic Assessment of Design Stability. PhD thesis, Lancaster University (2009)
17. Official Web-site of MSDN, https://msdn.microsoft.com/en-us/library/ee658105.aspx
18. Clarket, S., et al.: Separating Concerns throughout the Development Lifecycle. In: Intl. Workshop on Aspect-Oriented Programming ECOOP (1999)

19. Apel, S.: The Role of Features and Aspects in Software Development. PhD thesis, Otto-von-Guericke University Magdeburg (2007)
20. Tkachuk, M., Nagornyi, K.: Towards Effectiveness Estimation of Post Object-oriented Technologies in Software Maintenance. J. Problems in Programming, vol. 2-3 (special issue), pp.252--260 (2010)
21. Taromirad M., Paige, M.: Agile Requirements Traceability Using Domain-Specific Modeling Languages. In: Extreme Modeling Workshop, pp. 45--50 (2012)
22. Tarr, P.L., et al.: N Degrees of Separation: Multi-Dimensional Separation of Concerns. In: Proceedings of the International Conference on Software Engineering (ICSE), ACM, Los Angeles, USA, pp. 107--119 (1999)
23. Official Web-site of System Thinking World community, http://www.systems-thinking.org/kmgmt/kmgmt.htm
24. Tkachuk M., Martinkus I.: Models and Tools for Multi-dimensional Approach to Requirements Behavior Analysis. In: H.C. Mayr et al. (eds.) UNISCON 2012, LNBIP vol. 137, pp. 191--198. Springer-Verlag, Heidelberg (2013)
25. Saaty, T.L.: Fundamentals of the Analytic Hierarchy Process. RWS Publications (2000)
26. Zadeh, L.A.: Fuzzy Sets. WorldSciBook (1976)
27. Garlan, D., Monroe, R., Wile, D.: ACME: An Architecture Description Interchange Language. In: Proceedings of CASCON'97, p.p. 169--183, Toronto, Canada (1997)
28. Official Web-site of CIDE-project, http://wwwiti.cs.uni-magdeburg.de/iti_db/research/cide/
29. Nagornyi, K.: Elaboration and Usage of Method for Post Object-oriented Technologies Effectiveness's Assessment. J. East-European on Advanced Technologies, vol. 63, p.p. 21--25 (in Russian) (2013)

Provably correct graph transformations with small-t$\mathcal{ALC}^\star$

Nadezhda Baklanova[2], Jon Haël Brenas[1], Rachid Echahed[1], Christian Percebois[2], Martin Strecker[2], Hanh Nhi Tran[2]

[1] CNRS and Université de Grenoble
[2] Université de Toulouse / IRIT

Abstract. We present a prototype for executing and verifying graph transformations. The transformations are written in a simple imperative programming language, and pre- and post-conditions as well as loop invariants are specified in the Description Logic $\mathcal{ALC}$ (whence the name of the tool). The programming language has a precisely defined operational semantics and a sound Hoare-style calculus. The tool consists of the following sub-components: a compiler to Java for executing the transformations; a verification condition generator; and a tableau prover for an extension of $\mathcal{ALC}$ capable of deciding the generated verification conditions. A description of these components and their interaction is the main purpose of this paper.

Keywords: Graph Transformations, Programming Language Semantics, Tableau Calculus, Description Logic
Key Terms: ModelBasedSoftwareDevelopmentMethodology, FormalMethod, MathematicalModel, VerificationProcess

1 Introduction

Provably correct transformations of graph structures become increasingly important, for example for pointer manipulating programs, model driven engineering (such as EMF [1]) or the Semantic Web (with representation formats such as RDF [2]).

Contributions: This paper presents a new language, called small-t$\mathcal{ALC}$, and accompanying programming environment for executing graph transformations and reasoning about them. Let us characterize in a few words what our work is about and what it is not about:

– The primary aim of our development is to be able to *reason about graph transformations* in a pre- / post-condition style: can we ensure that any graph satisfying the pre-condition is transformed into a graph satisfying the post-condition? Essential ingredients of such a setup are a language for describing the transformations, and an assertional formalism for specifying the pre- and post-conditions.

⋆ Part of this research has been supported by the *Climt* project (ANR-11-BS02-016).

– The *transformation language* is an imperative programming language with special operations for manipulating graphs. This language is endowed with traditional control flow constructs (selection and loops) and elementary statements for adding and deleting arcs of a graph. There is a `select` statement that can be understood as a generalized, non-deterministic assignment operation and whose purpose is to perform matchings of rules in a target graph. After a high-level overview of small-t$\mathcal{ALC}$ (Section 2), we will give a more detailed account of the program logic (in Section 3.1) and transformation language (in Section 3.2). Our transformation language is by no means a full-fledged programming language: for example, arithmetic operations are excluded.
– The transformation language is *not graphical*, but textual. We do not question the utility and appeal of a graphical notation, but this issue is orthogonal to our concerns. We can imagine to couple small-t$\mathcal{ALC}$ with existing graphical editors, such as Henshin [3], in the sense of translating a graphical description of a rule to our textual format. The usefulness of the inverse direction is less evident, because the textual format is more expressive (offering, among others, nested loops and branching statements).
– The transformation language is *executable*, by a translation to Java (see Section 4): a code generator translates small-t$\mathcal{ALC}$ to Java code, which can then transform graphs specified in an appropriate format.

Altogether, we are thus primarily interested in *proofs of correctness* of graph transformations, for which two major approaches have emerged:

1. Model checking of graph transformations: given an initial graph and a set of transformation rules, check whether the graph can eventually evolve into a graph having certain properties, or whether specific properties can be ascertained to be always satisfied. This kind of reasoning is possible in principle (the initial graph can be specified by a pre-condition, invariants can be specified as loop conditions, eventuality properties as post-conditions), but our approach is clearly not geared towards this activity.
2. Full correctness proofs: given an arbitrary graph satisfying the pre-condition, verify that it evolves into a graph satisfying the post-condition. This is the kind of verification we are aiming at.

Full correctness proofs are hard, and undecidability of the generated proof obligations is a major concern for rich logics [4]. We propose to use a relatively simple logic, $\mathcal{ALC}$, belonging to the family of Description Logics (DLs). We summarize the logic in Section 3.1, and the fine-tuned interplay of the logic and the transformation language (among others: branching and loop conditions are formulas of this logic) brings it about that the proof obligations extracted from programs are decidable, as argued in Section 5. We are currently working on extending this approach to more expressive description logics, with the purpose of being able to tackle realistic problems in the areas of UML-style model transformations and RDF graph database transformations.

The work described here has reached the state of a sound prototype. In the corresponding sections, we will make precise which parts of the development are completed to which degree, and indicate which missing parts still have to be filled in. The small-t$\mathcal{ALC}$ environment is available from the following web page, where it will be regularly updated: http://www.irit.fr/~Martin.Strecker/CLIMT/Software/smalltalc.html.

Related work: Hoare-like logics have already been used to reason on graph transformations (see, e.g. [5]) but, as far as we are aware, no tool has been implemented. small-t$\mathcal{ALC}$, which is also based on a Hoare-like calculus, allows one to decide the verification problem, of programs operating on graphs, when the properties are expressed in the $\mathcal{ALC}$ logic. Some implementations of verification environments for pointer manipulating programs exist [6], however they often impose severe restrictions on the kind of graphs that can be manipulated, such as having a clearly identified spanning tree.

Other tools dedicated to reasoning on graph transformations have been proposed. For example, the GROOVE [7] system implements model-checking techniques using LTL or CTL formulas and thus departs from small-t$\mathcal{ALC}$ techniques.

The computation of weakest preconditions from a graph rewriting system is described by Habel, Pennemann and Rensink [8,9]. This work is concerned with extraction of weakest preconditions, but no proof system for the formulas is given. Pennemann [10] then describes a method of translating the extracted formulas to a resolution theorem prover. Radke [11] uses a more expressive logic: MSO. The spirit of the work described in this paper is similar, but we explicitly restrict the expressiveness of the logical framework to obtain decidable proof problems.

In a similar vein, Asztalos *et al.* [12] describe the verification of graph transformations based on category-theoretic notions and by translation to a logic for which no complete calculus is provided.

Raven[3] is a tool suite designed to handle and manipulate graph automata. In some sense Raven tends to generalize model-checking techniques from word to graph processing. Therefore techniques behind Raven tool are not directly comparable to small-t$\mathcal{ALC}$.

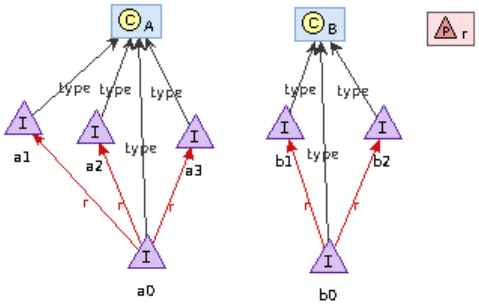
Alloy [13] is a popular framework for specifying and exploring relational structures, and it has been used to analyze graph transformations [14] written in the AGG transformation engine. Alloy interfaces with model checkers and can display counter models in case a transformation does not satisfy its specification. For verification, Alloy uses bounded model checking: errors for graphs of a certain size are systematically detected, but has the disadvantage that graphs beyond that size are not covered. As opposed to this, the proof method presented here is exhaustive, being based on a complete, decidable calculus.

[3] http://www.ti.inf.uni-due.de/research/tools/raven

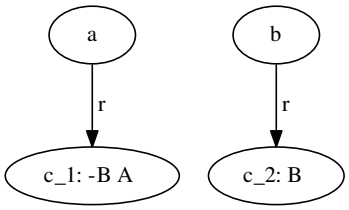
2 System Description

2.1 User's View

To explore the perspective of a user of small-t$\mathcal{ALC}$, we will walk through processing a simple program, but before, let us take a look at the kind of graphs we will be transforming, such as the example graph in Figure 1a (displayed with RDF-Gravity[4]). We will be processing graphs in RDF [2] format. These graphs consist of nodes and typed edges. The graphs are simple: there cannot be multiple edges of the same type between two nodes, but several edges, each of different type. In the example, there is only one type of relation (also called *role*): `r`. Here, instance node `a0` is linked with nodes `a1, a2` and `a3`; similarly `b0` with `b1` and `b2`. Nodes can be typed. In our example, we have two types (also called *concepts*) `A` and `B`. Nodes $\mathtt{a}_i$ are of type `A`, and nodes $\mathtt{b}_j$ of type `B`. It is a matter of display to represent concepts as (meta-)nodes in Figure 1a, and also the (meta-)relation `type` as arc linking a node to its type, but these meta-entities are subject to a different treatment than object nodes and relations.



(a) Input graph



(b) Counter model after failed proof

Fig. 1: Graph structures

Let us now turn to transformation programs, as the one depicted in Figure 2. A program is composed of one or several parameterized rules; and a parameterless `main` rule whose purpose is to specify the input- and output graph to be transformed and to identify the root nodes of the input graph. Rules can be assimilated to non-recursive procedures or macros. Procedural abstraction is so far not fully developed in our framework, so the analysis presented in the following concentrates on rule bodies.

The rule `ex_rule` has a precondition (`pre`) saying that node `a` is only connected (via arcs of type `r`) to nodes of type `A`, and that `b` is only connected to

[4] http://semweb.salzburgresearch.at/apps/rdf-gravity/

```
concepts A, B; roles r;

rule ex_rule (a, b) {
vars c;
pre: (a : ([!] r A)) && (b : ([!] r B));
select c with (b r c);
add(a r c);
post: (a : ([?] r B));
}
```

```
rule main () {
vars a, b;
ingraph "input_graph.rdf";
outgraph "output_graph.rdf";
a := node("a0");
b := node("b0");
ex_rule(a, b);
}
```

Fig. 2: An example program

nodes of type `B`. The program now does the following: among the nodes that `b` is connected to, we non-deterministically pick a node `c` and introduce an arc `r` between `a` and `c`. For example, the program might introduce an arc between `a0` and `b1` in the graph of Figure 1a (or between `a0` and `b2`). We can now assert that after running this program, the node that variable `a` points to is connected via `r` to at least one element of type `B`, as expressed in the postcondition.

Suppose the example program is in file `example.trans`. Running the verifier as follows confirms that the program is correct, *i.e.* that any graph satisfying the precondition is transformed into a graph satisfying the postcondition.

```
> graphprover example
starting proof ...formula valid
```

Let us modify the post-condition, claiming that `a` is exclusively connected to elements of type `B`: `post: (a : ([!] r B));`

When running the verifier again, we see that the property is incorrect, and that a counter-model has been created (see Figure 1b, here displayed with Graphviz[5]). This counter-model describes the state at the beginning of the program, namely a graph with four nodes, where c_1 is of type A and not of type B, and c_2 of type B. Clearly, when connecting a with c_2, the post-condition is violated.

We correct the post-condition, saying that `a` is only connected to elements of type `A` or `B`: `post: (a : ([!] r (A [||] B)));` Running the verifier again convinces us that this property is satisfied.

How does the verifier validate or invalidate a program? The approach is classic: from the annotated program, we extract a proof obligation by computing weakest pre-conditions (see Section 3.2). This is an $\mathcal{ALC}$ formula that is sent to a tableau decision procedure (described in Section 5.2). A failed proof attempt produces a saturated tableau from which a counter-model can always be extracted.

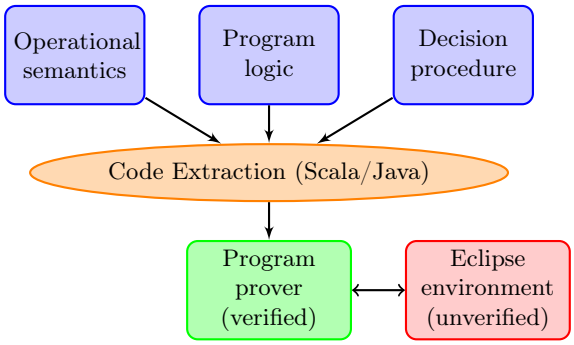


Fig. 3: Schema of formal development

2.2 Developer's View

Major parts of small-t$\mathcal{ALC}$ have a strong formal basis and are being developed in a proof assistant. We use Isabelle [15], but the formalization is easily adaptable to related proof assistants. Essential ingredients (see Figure 3) are the formalization of the program logic, the semantics of the programming language and a decision procedure of the extension of $\mathcal{ALC}$ we use (the latter has currently not been completely verified yet). This formalization (written in Isabelle's own functional and proof language) is automatically extracted to a general-purpose programming language, which is Scala in our case. We therefore obtain a highly reliable program prover, which is coupled with interface functionality (such as parsers and viewers) provided by Eclipse / Xtext to obtain the verifier described in Section 2.1. The transformation engine, described more in detail in Section 4, is so far unverified, but at least the Java code generator (Section 4) could be formally verified with by now standard compiler verification techniques.

3 Foundations

3.1 Logic

Our logic is a three-tier framework, the first level being Description Logic (DL) *concepts*, the second level *facts*, the third level *formulas* (Boolean combinations of facts and a simple form of quantification). Formulas occur not only in assertions (such as pre- and postconditions), but also in statements (Boolean conditions and `select` statement).

Concepts: In this paper, we concentrate on the description logic $\mathcal{ALC}$ [16]. For a being atomic concept names and r role (or relation) names, the abstract syntax of concepts C can be defined by the grammar:

[5] http://www.graphviz.org/

$$
\begin{array}{lll}
C ::= \bot & \text{(empty concept)} & \mid a \quad \text{(atomic concept)} \\
\mid \neg\, C & \text{(complement)} & \\
\mid C \sqcap C & \text{(intersection)} & \mid C \sqcup C \quad \text{(union)} \\
\mid ([?]\ r\ C) & \text{(some)} & \mid ([!]\ r\ C) \quad \text{(all)} \\
\mid C\tau & \text{(explicit substitution)} &
\end{array}
$$

The semantics of DLs is given by Kripke structures or, differently speaking, by typed graphs. Under this interpretation, *concepts* represent sets of individuals. The constructors $\neg, \sqcap, \sqcup$ (in Ascii notation: `!`, `[&&]`, `[||]`) then have the obvious meaning. $([?]\ r\ C)$ is the set of individuals x such that there is at least one r-typed edge $(x\ r\ y)$ between x and y, where y belongs to C. Dually, $([!]\ r\ C)$ is the set of individuals x all of whose r-edges go to individuals of type C.

The last constructor, *explicit substitution* [17], is a particularity of our framework, required for a gradual elimination of substitutions, as further described in Section 5.5. We have three kinds of substitutions τ:

- Replacement of a variable by another variable, of the form $[x := y]$,
- Adding a node v to / removing a node from an atomic concept a, of the form $[a := a + \{v\}]$ respectively $[a := a - \{v\}]$,
- Adding an edge (v_1, v_2) to / removing an edge from a role r, of the form $[r := r + \{(v_1, v_2)\}]$ respectively $[r := r - \{(v_1, v_2)\}]$.

Facts: Facts make assertions about an instance being an element of a concept, and about being in a relation. The grammar of facts is defined as follows:

$$
\begin{array}{lll}
fact ::= i : C & \text{(instance of concept)} \\
\mid i\ r\ i & \text{(instance of role)} \\
\mid i\ (\neg r)\ i & \text{(instance of role complement)} \\
\mid i \equiv i & \text{(equality of instances)} \\
\mid i \not\equiv i & \text{(inequality of instances)}
\end{array}
$$

Please note that since concepts are closed by complement, facts are closed by negation (the negation of a fact is again representable as a fact), and this is the main motivation for introducing the constructors "instance of role complement" and "inequality of instances".

Formulas: A formula is a Boolean combination of facts. We also allow quantification over individuals i (but not over relations or concepts), and, again, have a constructor for explicit substitution. We overload the notation $\bot$ for empty concepts and the Falsum.

$$
\begin{array}{lll}
form ::= \bot & \mid fact & \mid \neg form \\
\mid form \wedge form & \mid form \vee form & \\
\mid \forall i.form & \mid \exists i.form & \\
\mid form\ \tau & &
\end{array}
$$

In Figure 2, we use the Ascii notation `!`, `&&`, `||` for negation, conjunction and disjunction. The extension of interpretations from facts to formulas is standard. As usual, a formula that is true under all interpretations is called *valid*.

When calculating weakest preconditions (in Section 5.1), we obtain formulas which essentially contain no existential quantifiers; we keep them as constructor

because they can occur as intermediate result of computations. We say that a formula is *essentially universally quantified* if $\forall$ only occurs below an even and $\exists$ only below an odd number of negations. For example, $\neg(\exists x.\, x : C \wedge \neg(\forall y.\, y : D))$ is essentially universally quantified.

3.2 Programming Language

The programming language is an imperative language manipulating relational structures. Its distinctive features are conditions (in conditional statements and loops) that are restricted formulas of the logic $\mathcal{ALC}$, in the sense of Section 3.1. It has a non-deterministic assignment statement `select ... with` allowing to select an element satisfying a fact. Traditional types (numbers, arrays, inductive types) and accompanying operations are not provided; the language is thus only targeted at transformations of graphs.

Statements of our language are defined by the following grammar:

stmt ::=	`Skip`	(empty statement)
	`select` i `with` $form$	(assignment)
|	`delete(`i `:` C`)`	(delete element from concept)
|	`add(`i `:` C`)`	(add element to concept)
|	`delete(`i r i`)`	(delete edge from relation)
|	`add(`i r i`)`	(insert edge in relation)
|	*stmt* `;` *stmt*	(sequence)
|	`if` $form$ `then` *stmt* `else` *stmt*	
|	`while` $form$ `do` *stmt*	

Please note that the keywords `add` and `delete` are overloaded for nodes and for edges. There is no direct support for creating or deleting nodes in a graph, only for "moving" them between concepts. We intend to simulate node creation and deletion by providing a predefined concept `heap` such that `add(n: heap)` corresponds to creating node `n` and `delete(n: heap)` to deallocating node `n`. Details still have to be worked out.

The semantics is a big-step semantics with rules of the form $(st, \sigma) \Rightarrow \sigma'$ expressing that executing statement st in state σ produces a new state σ'.

The rules of the semantics are given in the Figure 4. Beware that we overload logical symbols such as $\exists$, $\wedge$ and $\neg$ for use in the meta-syntax and as constructors of $form$.

We do not enter into the details (also see the Isabelle formalization). Intuitively, the states σ manipulated by the operational semantics are the same as the interpretations of formulas, and they describe the current structure of a graph: which nodes are contained in each concept; which pair of nodes are contained in a role; and which variables are bound to which nodes. We write $\sigma(b)$ to evaluate the condition b (a formula) in state σ.

Most of the rules are standard, apart from the fact that we do not use expressions, but formulas as conditions. The auxiliary function $delete\_edge$ modifies the state σ by removing an r-edge between the elements represented by v_1 and v_2, and similarly for $generate\_edge$. There are analogous functions for adding / deleting in concepts.

$$(\textsc{Skip})\ \frac{}{(\texttt{Skip},\sigma)\Rightarrow\sigma} \qquad (\textsc{Seq})\ \frac{(c_1,\sigma)\Rightarrow\sigma'' \quad (c_2,\sigma'')\Rightarrow\sigma'}{(c_1;c_2,\sigma)\Rightarrow\sigma'}$$

$$(\textsc{EDel})\ \frac{\sigma' = delete\_edge\ v_1\ r\ v_2\ \sigma}{(\texttt{delete}(v_1\ r\ v_2),\sigma)\Rightarrow\sigma'} \qquad (\textsc{EGen})\ \frac{\sigma' = generate\_edge\ v_1\ r\ v_2\ \sigma}{(\texttt{add}(v_1\ r\ v_2),\sigma)\Rightarrow\sigma'}$$

$$(\textsc{SelAssT})\ \frac{\exists vi.(\sigma' = \sigma^{[v:=vi]} \wedge \sigma'(b))}{(\texttt{select}\ v\ \texttt{with}\ b,\sigma)\Rightarrow\sigma'}$$

$$(\textsc{IfT})\ \frac{\sigma(b) \quad (c_1,\sigma)\Rightarrow\sigma'}{(\texttt{if}\ b\ \texttt{then}\ c_1\ \texttt{else}\ c_2,\sigma)\Rightarrow\sigma'} \qquad (\textsc{IfF})\ \frac{\neg\sigma(b) \quad (c_2,\sigma)\Rightarrow\sigma'}{(\texttt{if}\ b\ \texttt{then}\ c_1\ \texttt{else}\ c_2,\sigma)\Rightarrow\sigma'}$$

$$(\textsc{WT})\ \frac{\sigma(b) \quad (c,\sigma)\Rightarrow\sigma'' \quad (\texttt{while}\ b\ \texttt{do}\ c,\sigma'')\Rightarrow\sigma'}{(\texttt{while}\ b\ \texttt{do}\ c,\sigma)\Rightarrow\sigma'} \qquad (\textsc{WF})\ \frac{\neg\sigma(b)}{(\texttt{while}\ b\ \texttt{do}\ c,\sigma)\Rightarrow\sigma}$$

Fig. 4: Big-step semantics rules

The statement `select` v `with` $F(v)$ selects an element vi that satisfies formula F, and assigns it to v. For example, `select` a `with` $a : A \wedge (a\ r\ b)$ selects an element a which is and instance of concept A and being r-related with a given element b.

`select` is a generalization of a traditional assignment statement. There may be several instances that satisfy F, and the expressiveness of the logic might not suffice to distinguish them. In this case, any such element is selected, non-deterministically. Let us spell out the precondition of (SelAssT): Here, $\sigma^{[v:=vi]}$ is an interpretation update for individuals, modifying σ for variable v and assigning it a value vi in the semantic domain. We check whether the formula b would be satisfied under this choice, and if it is the case, keep this assignment. In case no satisfying instance exists, the semantics blocks, *i.e.* the given state does not have a successor state, which can be considered as an error situation.

4 Executing Graph Transformations

Generating Java Code: For processing small-t$\mathcal{ALC}$ programs such as the one in Figure 2 and generating Java code, we use the Eclipse environment and, in particular, the Xtext[6] facilities for parsing, syntax highlighting and context-dependent help. The program prover is currently not fully integrated in this framework, so

[6] http://www.eclipse.org/Xtext/

that the interaction with the prover is performed via shell commands as described in Section 2.1.

In order to generate Java code for small-t$\mathcal{ALC}$ programs, we parse the program and then traverse the syntax tree with Xtext/Xtend, issuing calls to appropriate Java functions that manipulate a graph (which is initially the input graph provided in the program's `main` rule). Here is a glimpse at the Xtend code snippet that translates statements, in particular the `add` statement for roles:

```
def statement(Stmt s){
      switch s{
        Add_stmt: add(s.lvar,s.role,s.rvar)
        ...
      }
}
def add(String lvar,String role,String rvar)'''
       «graph».insertEdge(«lvar»,«role»,«rvar»);'''
```

Thus, a small-t$\mathcal{ALC}$ program fragment `add(a r b);` is translated to a Java call `g.insertEdge(a, r, b);`, where the graph `g` is the current graph.

Transforming Graphs: Once a Java program has been generated for a given small-t$\mathcal{ALC}$ program, it can be compiled and linked with a library that provides graph manipulating functions such as the above-mentioned `insertEdge`. When executing this program, it remains to read an input file containing a graph description, to perform the transformation and to output the new graph. We represent graphs in the RDF [2] format. Parsing and printing of RDF files is based on the Apache Jena framework[7].

5 Reasoning about Graph Transformations

5.1 Weakest Preconditions

For proving program correctness, we use a standard approach in program verification. For proving that a program *prog* establishes the postcondition Q if started in a state satisfying the precondition P, we calculate the weakest precondition of *prog* with respect to Q and then show that P implies this weakest precondition.

The details are inspired by the description in [18]: we compute weakest preconditions *wp* (propagating post-conditions over statements and taking loop invariants for granted) and verification conditions *vc* that aim at verifying loop invariants. Both take a statement and a DL formula as argument and produce a DL formula. For this purpose, while loops have to be annotated with loop invariants, and the `while` constructor becomes: `while` $\{form\}$ $form$ `do` *stmt*. Here, the first formula (in braces) is the invariant, the second formula the termination condition. The two functions are defined by primitive recursion over statements, see Figure 5 for the definition of *wp* (and the Isabelle sources for *vc*).

[7] http://jena.apache.org/

$$
\begin{array}{l}
wp(\texttt{Skip},\ Q) = Q \\
wp(\texttt{delete}(v :\ C),\ Q) = Q[C := C - \{v\}] \\
wp(\texttt{add}(v :\ C),\ Q) = Q[C := C + \{v\}] \\
wp(\texttt{delete}(v_1\ r\ v_2),\ Q) = Q[r := r - (v_1, v_2)] \\
wp(\texttt{add}(v_1\ r\ v_2),\ Q) = Q[r := r + (v_1, v_2)] \\
wp(\texttt{select}\ v\ \texttt{with}\ b,\ Q) = \forall v.(b \longrightarrow Q) \\
wp(c_1; c_2,\ Q) = wp(c_1, wp(c_2,\ Q)) \\
wp(\texttt{if}\ b\ \texttt{then}\ c_1\ \texttt{else}\ c_2,\ Q) = ite(b, wp(c_1, Q), wp(c_2, Q)) \\
wp(\texttt{while}\{iv\}\ b\ \texttt{do}\ c,\ Q) = iv
\end{array}
$$

Fig. 5: Weakest preconditions and verification conditions

Without going further into program semantics issues, let us only state the following soundness result that relates the operational semantics and the functions wp and vc:

Theorem 1 (Soundness). *If $vc(c, Q)$ is valid and $(c, \sigma) \Rightarrow \sigma'$, then $\sigma(wp(c, Q))$ implies $\sigma'(Q)$.*

What is more relevant for our purposes is the structure of the formulas generated by wp and vc, because it has an impact on the decision procedure for the DL fragment under consideration here. Besides the notion of "essentially universally quantified" introduced in Section 3.1, we need the notion of *quantifier-free* formula: A formula not containing a quantifier. In extension, we say that a statement is quantifier-free if all of its formulas are quantifier-free.

By induction on c, one shows:

Lemma 1 (Universally quantified). *Let Q be essentially universally quantified and c be a quantifier-free statement. Then $wp(c, Q)$ and $vc(c, Q)$ are essentially universally quantified.*

There is one major problem with the definition of function wp: the substitutions, such as $C := C - \{v\}$ or $r := r - (v_1, v_2)$. When conceiving them as a meta-operations, as is usually done, we see that substitutions would yield syntactically ill-formed formulas. For example, reducing $([?]\ r\ C)[C := C - \{v\}]$ would give $([?]\ r\ (C - \{v\}))$, which is not a valid concept expression. There are two ways out of this difficulty: we could either relax our syntax and accept expressions of the form $([?]\ r\ (C - \{v\}))$. This would induce a rather heavy change on the logic. Alternatively, we can treat substitution as a constructor of our language. This is the approach we have adopted, and therefore, substitutions appear as syntactic elements in the definitions of Section 3.1. It remains to be seen (in Section 5.2) how substitutions can be dealt with by proof methods of $\mathcal{ALC}$.

5.2 Tableau Method

The core of the decision procedure for proving the verification conditions that are obtained as described in Section 5.1 is a tableau calculus which combines

the traditional logical rules of a tableau calculus [19] with rules for progressively eliminating the substitutions which are not part of the logic $\mathcal{ALC}$.

As a consequence, and departing again from common practice in the DL literature, our tableau procedure does not manipulate facts (in the sense of Section 3.1), but formulas, *i.e.* Boolean combinations of facts. This extension becomes necessary because elimination of substitutions generates complex formulas. These could in principle be directly decomposed into sub-tableaux, but such a procedure obscures both the presentation and the implementation.

Preprocessing: The tableau manipulates quantifier-free formulas in negation normal form (nnf).

The formulas obtained from function vc do possibly contain quantifiers, but as mentioned before, the formulas are essentially universally quantified. To get rid of these quantifiers, we therefore perform the following steps:

- We convert the entry formula f to a prenex normal form, *i.e.* a form $\forall x_1 \ldots x_n.b$ with quantifier-free body b.
- We drop the quantifier prefix; more precisely, we replace the bound variables $x_1 \ldots x_n$ in b by free variables. This transformation preserves validity.
- We start the tableau with $nnf(\neg b)$. The procedure is a satisfiability check that either produces an empty tableau (meaning that f is valid) or a model of $\neg b$ that is a counter-example of f.

In negation normal form, negations only occur in front of atomic concepts (of the form $\neg a$, where a is an atomic concept). This invariant is maintained throughout the tableau procedure.

5.3 Tableau Rules

In the following, we present a high-level description of the tableau procedure. (The reader consulting the Isabelle theories will notice that the formalization is on two levels: a set-based, relational version, aiming at proving essential properties such as soundness and completeness of the rules; and a list-based implementation. The formal proofs of these theories are not yet finalized.)

A tableau manipulates sets of branches (also called *abox*es - "assertional boxes" in DL terminology). Each branch Γ is a set of formulas. We first concentrate on a set of rules aiming at decomposing formulas on a single branch. They have the form $\Gamma \hookrightarrow \Gamma'$, expressing that branch Γ is rewriten to Γ'. We write Γ, f instead of $\Gamma \cup \{f\}$ for adding formula f to Γ. The rules are displayed in Figure 6.

Let us comment on the rules: The structural rules conjC, disjCr, disjCl (for concepts) and conjF, disjFr, disjFl (for formulas) should be clear. The rule all allows to conclude $y : C$ if x is only r-connected to elements of type C, and there is an arc $(x\ r\ y)$. The rule some inserts an arc $(x\ r\ z)$ and a membership $z : C$ for an arbitrary z if it is known that x is r-connected to at least one element of type C. The rule eq propagates an equality $x \equiv y$ in the branch, provided the equality is not $x \equiv x$.

$$\textsc{conjC}\ \frac{(x : (C_1 \sqcap C_2)) \in \Gamma \quad \text{not}((x : C_1) \in \Gamma \text{ and } (x : C_2) \in \Gamma)}{\Gamma \hookrightarrow \Gamma, (x : C_1), (x : C_2)}$$

$$\textsc{disjCr}\ \frac{(x : (C_1 \sqcup C_2)) \in \Gamma \quad (x : C_1) \notin \Gamma \quad (x : C_2) \notin \Gamma}{\Gamma \hookrightarrow \Gamma, (x : C_1)}$$

$$\textsc{disjCl}\ \frac{(x : (C_1 \sqcup C_2)) \in \Gamma \quad (x : C_1) \notin \Gamma \quad (x : C_2) \notin \Gamma}{\Gamma \hookrightarrow \Gamma, (x : C_2)}$$

$$\textsc{all}\ \frac{(x : ([!]\ r\ C)) \in \Gamma \qquad (x\ r\ y) \in \Gamma \qquad (y : C) \notin \Gamma}{\Gamma \hookrightarrow \Gamma, (y : C)}$$

$$\textsc{some}\ \frac{(x : ([?]\ r\ C)) \in \Gamma \quad \text{for all } y, \text{not}((x\ r\ y) \in \Gamma \text{ and } (y : C) \in \Gamma)}{\Gamma \hookrightarrow \Gamma, (x\ r\ z), (z : C)}$$

$$\textsc{subst}\ \frac{(x : (C\tau)) \in \Gamma \quad nnf(push((x : C)\tau)) \notin \Gamma}{\Gamma \hookrightarrow \Gamma, nnf(push((x : C)\tau))}$$

$$\textsc{eq}\ \frac{(x \equiv y) \in \Gamma \quad x \neq y}{\Gamma \hookrightarrow \Gamma[x := y]}$$

$$\textsc{conjF}\ \frac{f_1 \wedge f_2 \in \Gamma \quad \text{not}(f_1 \in \Gamma \text{ and } f_2 \in \Gamma)}{\Gamma \hookrightarrow \Gamma, f_1, f_2}$$

$$\textsc{disjFr}\ \frac{f_1 \vee f_2 \in \Gamma \quad f_1 \notin \Gamma \quad f_2 \notin \Gamma}{\Gamma \hookrightarrow \Gamma, f_1} \qquad \textsc{disjFl}\ \frac{f_1 \vee f_2 \in \Gamma \quad f_1 \notin \Gamma \quad f_2 \notin \Gamma}{\Gamma \hookrightarrow \Gamma, f_2}$$

Fig. 6: Tableau rules

The rule SUBST is applicable for concepts with substitutions. As motivated in Section 5.1, substitutions cannot be eliminated at once, but they can be removed progressively, whenever the tableau prover hits on a fact of the form $(x : C\tau)$. Note that the variable x was possibly not present in the original tableau with which we have started the proof, but may have been introduced by a SOME-rule. If we encounter such a situation, we push the substitution as far as possible. We postpone the details to Section 5.5.

A branch Γ contains a *clash* ($clash(\Gamma)$) if either of the following holds:

- for x a variable, $(x : \bot) \in \Gamma$
- for x a variable and a an atomic concept, $(x : a) \in \Gamma$ and $(x : \neg a) \in \Gamma$
- for x, y variables, $(x\ r\ y) \in \Gamma$ and $(x\ (\neg r)\ y) \in \Gamma$

– for x a variable, $(x \not\equiv x) \in \Gamma$
– $\bot \in \Gamma$

5.4 Tableau Procedure

We can now formulate a depth-first-search function dfs exploring a tableau. The function takes a tableau (here implemented as a list of branches) and returns a list of models. Initially, the tableau is just the formula $[\{f\}]$ to be proved. If the resulting list is empty, f is not satisfiable. Otherwise, the list contains an element which is a model of f.

$$\begin{array}{lll} dfs[\,] & = [\,] & \\ dfs(\Gamma :: \Gamma_s) & = \texttt{if} & clash(\Gamma) \\ & \quad \texttt{then} & dfs(\Gamma_s) \\ & \quad \texttt{else} & \texttt{if}\ reducible(\Gamma) \\ & & \quad \texttt{then}\ dfs(\{\Gamma' | \Gamma \hookrightarrow \Gamma'\} @ \Gamma_s) \\ & & \quad \texttt{else}\ [\Gamma] \end{array}$$

The procedure progressively eliminates all inconsistent branches (with $clash(\Gamma)$). If a branch Γ is not inconsistent, but reducible (*i.e.* , there exists a Γ' with $\Gamma \hookrightarrow \Gamma'$), then we expand the tableau and explore the new branches.

5.5 Eliminating Substitutions

The *push* function used in the *subst* rule of Figure 6 pushes substitutions into formulas, "as far as possible". The remaining tableau rules then decompose formulas until substitutions hidden in subformulas become apparent and the *subst* rule can be applied again. Intuitively speaking, this process decreases the "height" of the substitutions in a formula, until they eventually disappear.

For a formula f, we define $push(f)$ as the formula f' which is the result of the rewrite system spelled out in the following. Thus: $push(f) = f'$ iff $f \rightsquigarrow^* f'$, where the rewrite relation $\rightsquigarrow$ is defined in the following. There are numerous cases to consider, and we do not present all of them.

Substitution in formulas are pushed into subformulas:

– $\bot\tau \rightsquigarrow \bot$
– $(\neg f)\tau \rightsquigarrow (\neg f\tau)$
– $(f_1 \wedge f_2)\tau \rightsquigarrow (f_1\tau \wedge f_2\tau)$
– $(f_1 \vee f_2)\tau \rightsquigarrow (f_1\tau \vee f_2\tau)$

Substitution in facts: Substitutions of individual variables $f[x := y]$ are carried out as expected. Otherwise, we procede as follows:

– $(x : \neg C)\tau \rightsquigarrow x : (\neg C\tau)$
– $(x : C_1 \sqcap C_2)\tau \rightsquigarrow x : (C_1\tau \sqcap C_2\tau)$
– $(x : C_1 \sqcup C_2)\tau \rightsquigarrow x : (C_1\tau \sqcup C_2\tau)$

- For substitutions τ of the form $a := a - \{v\}$ or $a := a + \{v\}$:
  - $(x : c)[a := a - \{v\}] \rightsquigarrow (x : c)$ for $a \neq c$, and similarly for $a := a + \{v\}$
  - $(x : a)[a := a - \{v\}] \rightsquigarrow (x : a) \wedge x \neq v$
  - $(x : a)[a := a + \{v\}] \rightsquigarrow (x : a) \vee x = v$
  - $(x : ([?]\ r\ C))[a := a - \{v\}] \rightsquigarrow (x : ([?]\ r\ C[a := a - \{v\}]))$, and similarly for the other combinations involving constructor $[?]$ or $[!]$ and substitutions $a := a + / - \{v\}$.
- For substitutions τ of the form $r := r - \{(v_1, v_2)\}$ or $r := r + \{(v_1, v_2)\}$:
  - $(x : c)[r := r - \{(v_1, v_2)\}] \rightsquigarrow x : c$, and similarly for $r + \{(v_1, v_2)\}$
  - $(x : ([!]\ r'\ C))[r := r - \{(v_1, v_2)\}] \rightsquigarrow (x : ([!]\ r'\ C))$ for $r \neq r'$
  - $(x : ([!]\ r\ C))[r := r - \{(v_1, v_2)\}] \rightsquigarrow$

    $$\begin{array}{l} ite\ ((x = v_1) \wedge (v_2 : (\neg C[r := r - (v_1, v_2)])) \wedge (v_1\ r\ v_2), \\ \quad (x : (<\ 2\ r\ (\neg C[r := r - (v_1, v_2)]))), \\ \quad (x : ([!]\ r\ C[r := r - (v_1, v_2)]))) \end{array}$$

    Here, *ite* is for if-then-else: $ite(a, b, c) = (a \longrightarrow b) \wedge (\neg a \longrightarrow c)$.
    Please note that the logic $\mathcal{ALC}$ cannot completely express the effect of substitution, and we have to resort to the more expressive logic $\mathcal{ALCQ}$, which turns out to be complete for substitutions. Thus, the "then" branch of the *ite* construct expresses that x is r-connected to less than 2 elements of $(\neg C[r := r - (v_1, v_2)])$. We have however not yet implemented tableau rules for $\mathcal{ALCQ}$, so we stick to the simpler logic in this presentation.
  - $(x : ([!]\ r\ C))[r := r + \{(v_1, v_2)] \rightsquigarrow$
    $\neg((x = v_1) \wedge (v_2 : \neg(C[r := r + (v_1, v_2)])) \wedge (v_1\ (\neg r)\ v_2))$
    $\wedge(x : ([!]\ r\ (C[r := r + (v_1, v_2)])))$
  - Similar rules for existential quantification $(x : ([?]\ r\ C))$.

6 Conclusions

We have presented small-t$\mathcal{ALC}$, a framework for executing graph transformations and proving their correctness with a sound and complete calculus. One of the distinctive features of the approach is its formal semantic basis. We are now moving towards application, such as Sparql Query and Update in the knowledge representation world, and model transformations as used in model-driven engineering. The greatest challenge is the development of logics that are more expressive than $\mathcal{ALC}$ but remain decidable. Even though a low proof-theoretic complexity is not a major concern for program correctness proofs (these are not executed on a large knowledge base), the concern changes when wanting to execute programs efficiently on a large data set.

Acknowledgements We are grateful to María Espinoza who has helped us explore the applicability of graph transformations to the RDF world [20].

References

1. Budinsky, F., Brodsky, S.A., Merks, E.: Eclipse Modeling Framework. Pearson Education (2003)
2. Cyganiak, R., Lanthaler, M., Wood, D.: RDF 1.1 Concepts and Abstract Syntax. http://www.w3.org/TR/rdf11-concepts (2014)
3. Arendt, T., Biermann, E., Jurack, S., Krause, C., Taentzer, G.: Henshin: Advanced concepts and tools for in-place EMF model transformations. In: Proceedings of MoDELS'10. Volume 6394 of LNCS. Springer (2010)
4. Immerman, N., Rabinovich, A., Reps, T., Sagiv, M., Yorsh, G.: The boundary between decidability and undecidability for transitive-closure logics. In Marcinkowski, J., Tarlecki, A., eds.: Computer Science Logic. Volume 3210 of LNCS. Springer Berlin / Heidelberg (2004) 160–174
5. Poskitt, C.M., Plump, D.: Hoare-style verification of graph programs. Fundamenta Informaticae **118**(1-2) (2012) 135–175
6. Møller, A., Schwartzbach, M.I.: The pointer assertion logic engine. In: PLDI. (2001) 221–231
7. Ghamarian, A.H., de Mol, M., Rensink, A., Zambon, E., Zimakova, M.: Modelling and analysis using GROOVE. STTT **14**(1) (2012) 15–40
8. Habel, A., Pennemann, K.H., Rensink, A.: Weakest preconditions for high-level programs. In Corradini, A., Ehrig, H., Montanari, U., Ribeiro, L., Rozenberg, G., eds.: Graph Transformations (ICGT), Natal, Brazil. Volume 4178 of LNCS. Springer Verlag, Berlin (September 2006) 445–460
9. Habel, A., Pennemann, K.H.: Correctness of high-level transformation systems relative to nested conditions. MSCS **19**(02) (2009) 245–296
10. Pennemann, K.H.: Resolution-like theorem proving for high-level conditions. In Ehrig, H., Heckel, R., Rozenberg, G., Taentzer, G., eds.: Graph Transformations. Volume 5214 of LNCS. Springer Berlin / Heidelberg (2008) 289–304
11. Radke, H.: HR* graph conditions between counting monadic second-order and second-order graph formulas. ECEASST **61** (2013)
12. Asztalos, M., Lengyel, L., Levendovszky, T.: Formal specification and analysis of functional properties of graph rewriting-based model transformation. Software Testing, Verification and Reliability **23**(5) (2013) 405–435
13. Jackson, D.: Software Abstractions: Logic, language, and analysis. MIT Press (2012)
14. Baresi, L., Spoletini, P.: On the use of Alloy to analyze graph transformation systems. In Corradini, A., Ehrig, H., Montanari, U., Ribeiro, L., Rozenberg, G., eds.: Graph Transformations. Volume 4178 of LNCS. Springer (2006) 306–320
15. Nipkow, T., Paulson, L., Wenzel, M.: Isabelle/HOL. A Proof Assistant for Higher-Order Logic. Volume 2283 of LNCS. Springer Berlin / Heidelberg (2002)
16. Baader, F., Sattler, U.: Expressive number restrictions in description logics. Journal of Logic and Computation **9**(3) (1999) 319–350
17. Abadi, M., Cardelli, L., Curien, P.L., Lévy, J.J.: Explicit substitutions. Journal of Functional Programming **1**(4) (October 1991) 375–416
18. Nipkow, T., Klein, G.: Concrete Semantics. http://www21.in.tum.de/~nipkow/Concrete-Semantics/ (2014)
19. Baader, F., Sattler, U.: Tableau algorithms for description logics. In Dyckhoff, R., ed.: Automated Reasoning with Analytic Tableaux and Related Methods. Volume 1847 of LNCS. Springer (2000) 1–18
20. Espinoza, M.V.: Transformation de graphes en RDF. Master's thesis, Université de Toulouse (2014)

A Study of Bi-Objective Models for Decision Support in Software Development Process

Vira Liubchenko[1],

[1] Odessa National Polytechnic University, 1 Shevchenko av.,
65044 Odessa, Ukraine
lvv@edu.opu.ua

Abstract. This paper is concerned with the bi-objective problem in search-based software engineering for high-level decision-making. The paper presents bi-objective models for next release problem and modularization quality problem that characterized by the presence of two conflicting demands, for which the decision maker must find a suitable balance. The complex nature of such kind of problem has motivated the application of heuristic optimization techniques to obtain Pareto-optimal solutions. In this case, limitation on the size of the problem is reasonable.

Keywords. Search-based software engineering, bi-objective model, next release problem, modularization problem.

Key Terms. Model, mathematical model, software engineering process.

1 Introduction

Search-Based Software Engineering (SBSE) has become a subfield of software engineering characterized by growing of activity and research interest. SBSE seeks to reformulate Software Engineering problems as 'search problems' [1] in which optimal or near-optimal solutions are sought in a search space of candidate solutions, guided by a fitness function that distinguishes between better and worse solutions.

It has been argued that the virtual nature of software makes it well suited for Search-Based Optimization (SBO) [2]. This is because fitness is computed directly in terms of the engineering artifact, without the need for the simulation and modeling inherent in all other approaches to engineering optimization. This simplicity and ready applicability make SBSE a very attractive option.

Traditionally SBSE has based on finding the optimal or near-optimal solution to the problem with respect to a single objective. However, single-objective approach often is incorrect because of existing of many incomparable objectives in the

framework of one problem. Incomparability of objectives makes inapplicable waiting of the different objectives in order to combine them into a single weighted sum objective.

This reason has caused applying of multi-objective approaches in SBSE and using SBSE as a tool for decision support. To underpin the focus on decision support, SBO problem should be formulated as multi-objective problems, to which a Pareto optimal approach can be applied [3]. In Pareto optimal approaches, the outcome is a set of candidate solutions, each of which cannot be enhanced according to one of the multiple objectives to be optimized without a negative impact on another.

In this paper, we explore existing bi-objective approaches for high-level decision support in software development process. The rest of the paper is organized as follows. Section 2 briefly describes using of bi-objective models for Next Release and Modularization Problems. Section 3 presents SBO on decision-making perspective. Finally, section 4 draws the main conclusions.

2 Bi-Objective Models at the Early Stages of Software Development

Software engineers have been exploiting many different software development methodologies that recommend different framework of stages. In this paper, we base on the fact that high-level decision most often need support on requirement specification and design stages, which present, more or less, in every methodology. To explore bi-objective models at these stages, we use papers gathered in repository of publications on SBSE [4].

2.1 Requirement Specification Stage

One of the core problems of requirement specification stage in incremental methodologies is Next Release Problem (NRP). Decision maker determines which features should be included in the next release of the product in order to satisfy the highest possible number of customers and entail the minimum cost for the company [3]. NRP is a form of cost-benefit analysis for which a Pareto optimal approach is attractive.

In NRP a set of customers, $C = \{c_1, ..., c_m\}$, each customer has a degree of importance for the company that can be reflected by a weight factor, $Weight = \{w_1, ..., w_m\}$, where $w \in [0,1]$ and $\sum_{j=1}^{m} w_j = 1$.

It is assumed that there is the set of independent requirements, $R = \{r_1, ..., r_n\}$, that are targeted for the next release of an existing software system. Satisfying each requirement entails spending a certain amount of resources, which can be translated into cost terms, $Cost = \{cost_1, ..., cost_n\}$.

Satisfaction of requirements provides value for the company. The level of satisfaction for a given customer depends on the subset of requirements that are satisfied in the next release of the software product. The requirements are not equally important for a given customer. Each customer c_j ($1<j<m$) assigns a value to

requirement r_i ($1<i<n$) denoted by $value(r_i, c_j)$ where $value(r_i, c_j)>0$ if customer j has the requirement i and 0 otherwise.

In the formulation of the bi-objective NRP, two objectives are taken into consideration in order to maximize customer satisfaction (or the total value for the company) and minimize required cost. Let the decision vector $\vec{x} = \{x_1,\ldots,x_n\} \in \{0,1\}$ determines the requirements that are to be satisfied in the next release. In this vector, x_i is 1 if the requirement i is selected and 0 otherwise.

The first objective function is considered for maximizing total value:

$$\text{Maximize} \sum_{i=1}^{n} x_i \sum_{j=1}^{m} w_j \cdot value(r_i, c_j).$$

The problem is to select a subset of the customers' requirements, which results in the maximum value for the company.

The second objective function is considered for minimizing total cost required for the satisfaction of customer requirements:

$$\text{Minimize} \sum_{i=1}^{n} cost_i \cdot x_i .$$

In order to convert the second objective to a maximization problem, the total cost is multiplied by -1. Therefore, the bi-objective model can be represented as follows:

$$\begin{aligned} &\text{Maximize } f_1(\vec{x}) = \sum_{i=1}^{n} x_i \sum_{j=1}^{m} w_j \cdot value(r_i, c_j) \\ &\text{Maximize } f_2(\vec{x}) = -\sum_{i=1}^{n} cost_i \cdot x_i \end{aligned} \quad . \qquad (1)$$

2.2 Design Phase

Software design usually includes low-level component and algorithm design and high-level, architecture design. A high-level software engineering problem that is most related to software architectures is Modularization Problem (MP). Decision Maker finds the best grouping of components to subsystems. For that, structure of software system is transformed into a directed graph G, the main question to be answered is what constitutes a good partition of the software structure graph. The goodness of a partition is usually measured with a combination of cohesion and coupling.

Cohesion is a measure of the degree to which the components of a single subsystem belong together. A high cohesion indicates a good modularization arrangement because the components grouped within the same subsystem are highly dependent on each other. A low cohesion, on the other hand, generally indicates a

poor modularization arrangement because the components grouped within a subsystem are not strongly related.

The cohesion A_i of subsystem i with N_i components is defined as:

$$A_i = \frac{\mu_i}{N_i^2},$$

where μ_i is the number of intra-edge dependencies (relationships to and from components within the same subsystem), N_i^2 is the maximum number of possible dependencies between the components of subsystem i.

Coupling is a measure of the connectivity between distinct subsystems. A high degree of coupling is undesirable because it indicates that subsystems are highly dependent on each other. Conversely, a low degree of coupling is desirable because it indicates that individual subsystems are largely independent of each other.

The coupling E_{ij} between subsystems i and j, each consisting of N_i and N_j components respectively, is defined as:

$$E_{ij} = \begin{cases} 0 & \text{if } i = j \\ \dfrac{\varepsilon_{ij}}{2N_iN_j} & \text{if } i \neq j \end{cases},$$

where ε_{ij} is the number of inter-edge dependencies (relationships to and from components of subsystems i and j).

In the formulation of the bi-objective MP, two objectives expresses the tradeoff between cohesion and coupling are taken into consideration in order to create highly cohesive subsystems and penalize the creation of too many dependencies between subsystems.

Given software structure graph G partitioned into k clusters, modeled partition of software system into subsystems, we define MP as:

$$\begin{aligned} &\text{Maximize } f_1(\vec{x}) = \frac{1}{k}\sum_{i=1}^{k} A_i \\ &\text{Maximize } f_2(\vec{x}) = -\frac{k(k-1)}{2}\sum_{i=1}^{n}\sum_{j=1}^{k} E_{ij} \end{aligned}. \qquad (2)$$

3 SBO as Decision Support

SBO can be applied to situations in which the human will decide on the solution to be adopted, but the search process can provide insight to help guide the decision maker. This insight agenda, in which SBO is used to gain insights and to provide decision support to the software engineering decision maker, has found natural resonance and

applicability when used at the early stages of the software engineering lifecycle, where the high-level decisions made can have far-reaching implications.

Many of the values used to define a problem for optimization, particularly at the early stages of the software development process, come from estimates. In these situations, it is not optimal solutions that the decision maker requires, as much as guidance on which of the estimates are most likely to affect the solutions. Therefore, SBO is not merely a research program in which one seeks to 'solve' software engineering problems; it is a rich source of insight and decision support.

Bi-objective problems stated above are NP-hard, and, therefore, cannot be solved using exact optimization techniques for large-scale problem instances. That is why metaheuristic search techniques are usually applied to find approximations of Pareto optimal set (or front) for the bi-objective problem. Decision maker selects the solution from the found set according to his (her) preferences.

Restriction on SBO approach connected with the point at which the problem becomes too small. For NRP, limitation is a function of the number of requirements, which should exceed about 20 requirements. By contrast, there is no number of customers that is too small for the problem to be worthwhile. For MP, limitation is a function of the number of components, which should exceed about 20 components.

4 Conclusion

Vital errors in software engineering such as too many requirements being realized in release and poor quality of software architecture are caused by false intuition of the decision maker. SBO can address this problem, it automatically scour the search space for the solutions that best fit the human assumptions in the objective functions. However, it has been widely observed that search techniques are good at producing unexpected answers. Automated search techniques effectively work in tandem with the human encapsulating human assumptions and intuition.

Future work will consider modification of SBO for including dependency relationship between requirements in NRP, between components in MP, and exploring the integrated model for both problems.

References

1. Harman, M., Jones, B.F.: Search based software engineering. Information and Software Technology, 43(14), pp. 833--839 (2001)
2. Harman, M.: Why the virtual nature of software makes it ideal for search based optimization. In: Proceedings of the 13th International Conference on Fundamental Approaches to Software Engineering (FASE'10). LNCS, vol. 6013, pp. 1--12. Springer, Heidelberg (2010)
3. Durillo, J.J., Zhang, Y., Alba, E., Harman, M., Nebro, A.J.: A study of the bi-objective next release problem. Empirical Software Engineering, vol. 16(1), pp. 29--60 (2011)
4. Repository of Publications on Search Based Software Engineering, http://crestweb.cs.ucl.ac.uk/resources/sbse_repository/repository.html

5. Doval, D., Mancoridis, S., Mitchell, B.S.: Automatic Clustering of Software Systems using a Genetic Algorithm. In: Proceedings of Software Technology and Engineering Practice, pp. 73--91 (1998)

Method of Evaluating the Success of Software Project Implementation Based on Analysis of Specification Using Neuronet Information Technologies

Tetiana Hovorushchenko[1], Andriy Krasiy[2]

[1] Khmelnitsky National University, Khmelnitsky, Ukraine
tat_yana@ukr.net
[2] Khmelnitsky National University, Khmelnitsky, Ukraine
andriy-krasiy@yandex.ua

Abstract. The actuality and importance of skill to evaluate the possible success of software project based on SRS were showed in this paper. The aim of research is prediction of characteristics and evaluating the success of software project implementation based on analysis of SRS. Method of evaluating the success of software project implementation based on analysis of SRS using neuronet information technologies was first proposed. This method provides the prediction of success of software projects implementation, comparison of software projects on the basis of SRS and choice of the best SRS of project.

Keywords: software requirements specification (SRS), software project, success of project implementation, SRS indicators, project characteristics, integrative indicator of project, the degree of success of the project implementation.

Key Terms: Model-Based Software System Development, Software Component, Software System, Specification Process.

1 Introduction

Statistics of success of software projects implementation according to The Standish Group International [1] showed that the rate of challenged projects (that late, over budget, and/or with less than the required features) is the constant value (42-46% projects). These statistics reflect the high rate of non-quality (the failed and the challenged) software projects in terms of interpretation of software quality [2].

As shown in [3], the errors of requirements formulation are 10-25% of all errors. The analysis of errors of embedded and application software, which were made at the stage of the requirements formulation, is given in [4]. In [5-7] the fact is confirmed, that the causes of many incidents and accidents through software are in the SRS, rather than in coding. In [6] the experiment is described, which showed that the software versions written by different developers for the same requirements, contain the joint errors associated with errors of SRS. These experimental statements leads to

the need to deepen of the SRS analysis. So *the actual and important* is the skill of evaluation of the success of project implementation on the basis of SRS. *The aim of this research* is the prediction of the characteristics and evaluation of success of implementation of software project based on the SRS analysis.

The success of software project implementation is timely execution of software project within the allocated budget and with realization of all necessary features and functionality. It can be estimated at the design stage based on the predicted values of the main project characteristics [8-10] - duration, cost, complexity, cross-platform, usability and quality. *Duration* is the sequence of the project stages based on the needs of project management. The relative duration is evaluated as compared to other software projects. *Cost* is difficult to assess at the early stages because it is highly dependent on the number of lines of code (the cost of one line is 0.5$). At the early stages of the life cycle we can evaluate the relative cost (as compared to other projects). *Complexity* is determined by the number of interacting components, the number of connections between the components and the complexity of their interactions. *Cross-platform* is the ability of software to run on more than one hardware platform and/or operating system. *Usability* is effectiveness, profitability and satisfaction of users by software project. *Quality* is the degree of compliance with the software characteristics of requirements. From the determinations of characteristics it is clear that none of them are part of other characteristic, that justifies this choice [9, 10].

Analysis shows, that the existing methods and tools [9, 10] of characteristics determination are not suitable to evaluation of their values at the stage of requirements formulation, since they focus on the ready source code. The known methods (Using natural language processing technique, Using CASE analysis method, QAW-method, Using global analysis method, O'Brien's approach, Method to discover missing requirement elicitation, Selection of elicitation technique, Comparison and categorization of requirements elicitation techniques, Techniques for ranking and prioritization of software requirements) and tools (OSRMT, Tools by LDRA, Sigma Software, DEVPROM, CASE.Analytics) of SRS analysis and existing technologies of risk management (SEI, SRE, CRM, TRM, FSI, ERM) [9-13] are not suitable for quantitative evaluation of the project characteristics, because all are targeted to control over compliance with requirements of SRS, but none of them define the predicted values of characteristics on the SRS analysis.

Then for prediction of success of software project implementation on the analysis of SRS *the task of research* is development of method of evaluating the success of software project implementation based on analysis of specification.

2 Method of Evaluating the Success of Software Project Implementation Based on Analysis of Specification Using Neuronet Information Technologies (MESSPI)

Method of evaluating the success of software project implementation based on analysis of SRS consists of next stages: 1) neuronet prediction of characteristics of software project based on the analysis of specification; 2) interpretation of the received relative

values of the software project characteristics; 3) evaluation of the degree of success of the software project implementation; 4) testing of the stability and acceptability of compensations of software project characteristics.

Let the software project is specified by the SRS [14] in the next formalized form:

$$SRS=<R1,R2,R3,R4>, \quad (1)$$

where R1 – the set of indicators of section1 of the SRS, R2 – indicators of section2, R3 – indicators of section3, R4 – indicators of section4. Selection and possible values of SRS indicators from the sets R1-R4 were detailed in [9].

The *first stage of MESSPI* is prediction of software project characteristics on the SRS analysis, result of that is determining of the relative values of characteristics:

$$SCH=\{Cs,Dsp,Cx,Cp,Ub,Qs\}, \quad (2)$$

where Cs – software project cost, Dsp –duration, Cx –complexity, Cp – cross-platform, Ub – usability, Qs – quality.

Some indicators of specification [9] affect the above characteristics, but equations is not known, by which can calculate the characteristic value on the basis of the sets of SRS indicators – all available formulas of characteristics evaluation is oriented to ready source code [9, 10]. Hecht-Nielsen's theorem proves the possibility of solving the task of representation of multidimensional function of arbitrary form on the artificial neural network (ANN). Therefore, ANN will be used to implement of the unknown functions of dependence of the project characteristics on SRS indicators. In [9] the ANN was developed, which processes and approximates the set of SRS indicators and provides the predicted quantitative values of characteristics - Fig. 1. Selection and possible values of ANN inputs, equations for ANN functioning and forming of ANN outputs (predicted relative values of the characteristics) were detailed in [9], so this information is not represented in this paper.

ANN of characteristics prediction based on the SRS analysis was trained so that all values of characteristics are the values of the interval (0, 1]. The value of each characteristic nearly to 0 negative affects on the success of project implementation (high cost, duration and complexity; low quality, usability, cross-platform). The value nearly to 1 positive impacts on the success of the project implementation (low cost, duration, complexity; high quality, usability, cross-platform).

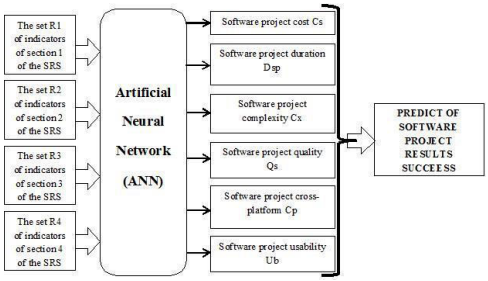


Fig. 1. The concept of neuronet prediction of characteristics of software project based on the analysis of specification

Let the ANN provided the following set of values of characteristics of project Sp:

$$SCH_{ANN}=\{Cs_{ANN}, Cx_{ANN}, Dsp_{ANN}, Ub_{ANN}, Cp_{ANN}, Qs_{ANN}\} \quad (3)$$

The developers and customers are difficult to comprehensively assess the success of software project implementation on the basis of the ANN's relative values of main characteristics. Therefore, the *second stage of MESSPI* is the interpretation of the received relative values of the project characteristics.

For this we introduce the integrative indicator of software project. *Integrative indicator* Iip_{Sp} – is the quantitative indicator of project implementation success based on the set SCH_{ANN}. We cannot to establish mutual dependence of them and to determine their impact on the integrative indicator of software project - these formulas and functions are not available. Therefore, we assume that all six predicted characteristics are equally important to the success of the project, and the integrative indicator of project depends equally on all six characteristics. In the absence of formulas and functions the simplest and the most obvious way of definition of integrative indicator of project is the using of its graphic presentation (in the classic radar chart, the axes of which there are six characteristics of the project - Fig. 2). Then the integrative indicator of project is area of figure, which are shaped the predicted (by ANN) values of the project characteristics. Because ANN predicts the values of 6 characteristics, the coordinate system (Radar chart) will have 6 axes (the angle between the axes is 60°), and in accordance the integrative indicator of project is area of the hexagon $Cs_{ANN}Cx_{ANN}Dsp_{ANN}Ub_{ANN}Cp_{ANN}Qs_{ANN}$ highlighted thick line on Fig. 3.

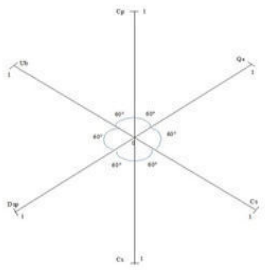


Fig. 2. The coordinate system for Iip_{Sp}

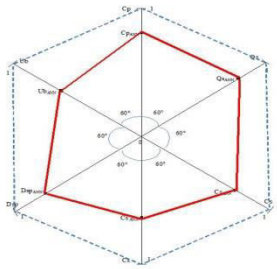


Fig. 3. The graphical representation of Iip_{Sp} and Iip_{max}

For calculation of integrative indicator Iip_{Sp} we will divide the hexagon into six triangles, will calculate the area of each triangle with two sides (value of characteristics) and angle between them (60°) and will add the obtained values of triangles areas:

$$S_{CsOCx}=½*Cs_{ANN}*Cx_{ANN}*\sin 60°=0.5*0.866*\ Cs_{ANN}*Cx_{ANN}, \quad (4)$$

$$Iip_{Sp}=0.5*0.866*(Cs_{ANN}*Cx_{ANN}+\ Cx_{ANN}*Dsp_{ANN}+\ Dsp_{ANN}*Ub_{ANN}+\ Ub_{ANN}*Cp_{ANN}+$$

$$+Cp_{ANN}*Qs_{ANN}+\ Qs_{ANN}*Cs_{ANN}) \quad (5)$$

The order of hexagon axes was selected taking into account of features of ANN training and for reasons of inability of compensation of the low values of some characteristics by high values of other characteristics (as all six characteristics are

important for the software project). Formula (5) shows that pairwise multiplication of the characteristics values can allow these compensations. Therefore, the upper part of the coordinate system has three axes for characteristics Ub, Cp, Qs, and the lower part consists of three axes for characteristics Dsp, Cx, Cs, for which the rule of ANN training is: the value of characteristic nearly to 0 means high cost, duration, complexity and low quality, usability, cross-platform. The junction of axes for characteristics from different categories was selected in pairs exactly as low value of cost (Cs→1) shall not compensate low value of quality (Qs→0), short value of duration (Dsp→1) can not compensate low value of usability (Ub→0).

We will need also the maximum possible value of integrative indicator of project: Iip_{max} – is the area of hexagon CsCxDspUbCpQs highlighted dotted line on Fig. 3. ANN was trained so that maximum possible value of each characteristic – is 1. Then:

$$Iip_{max}=0.5*0.866*(\ 1*1+\ 1*1+\ 1*1+\ 1*1+\ 1*1+\ 1*1)=2.598 \quad (6)$$

By itself, the integrative indicator of project is uninformative to the developer and customer due to the difficulty of interpretation of its value, therefore the *third stage of MESSPI* is the evaluation of the degree of success of project implementation based on the integrative indicator of project. The value $Iip_{max}=2.598$ – is the best value of integrative indicator, then the degree P_{Iip} of success of project implementation is:

$$P_{Iip}=Iip_{Sp}/Iip_{max}=Iip_{Sp}/2.598=0.385*Iip_{Sp} \quad (7)$$

The value of the degree of success of the software project implementation nearly to 0 indicates the low success of software project implementation.

As mentioned above, the compensation of values of the characteristics with the same value of integrative indicator is not always correct. Then the *fourth stage of MESSPI* is the testing of the stability and acceptability of characteristics compensations. If the hexagon $Cs_{ANN}Cx_{ANN}Dsp_{ANN}Ub_{ANN}Cp_{ANN}Qs_{ANN}$ (area of which is the integrative indicator) will be convex, the characteristics of software project is considered the stable, and their compensatory effects are acceptable (valid). We introduce the *indicator* Ace_{Sp} *of stability and acceptability of compensatory effects of the characteristics.* This indicator will take the value "True", if characteristics are stable, their compensatory effects are acceptable (i.e. hexagon is convex).

Criterion of convexity of hexagon is the simultaneous fulfillment of two conditions: 1) the same sign of sines of all angles of the hexagon; 2) the sum of all the angles of hexagon is 720° (by theorem about sum of the angles of convex polygon).

Here are the steps to determine of the angles of the hexagon (by Fig. 3): 1) calculate the unknown third side for each triangle by law of cosines; 2) find one unknown angles in each triangle by law of cosines; 3) find second unknown angle in each triangle by theorem about the sum of angles; 4) find the angles of the hexagon.

After finding of the angles of the hexagon we should find sines of obtained angles and compare their signs. And we should find the sum of the obtained angles and compare this sum with 720°. If the sum of the angles of hexagon is 720° and sines of angles have the same signs, then hexagon is convex, accordingly indicator of stability and acceptability of compensatory effects of the characteristics Ace_{Sp}=True.

3 Experiments

We performed experiments on the practical use of the MESSPI. For this we considered four alternative software projects, developed by different teams of developers to solve the same task – development of support system (web-portal) for practices of students of IT-specialties. Each development team consists of three IT professionals: project manager, requirements engineer and web-developer. Specialists from different teams had the same level of qualifications and the same experience in similar projects: project manager and requirements engineer of each team previously worked in three similar successful projects, web-developer of each team previously worked in two similar successful projects. All four development teams represented the different software companies of Khmelnitsky. Each development team had the equal opportunity to communicate with the customer for identification of customer requirements. Three joint meetings of all developers of four teams and representatives of the customer were organized. In addition, individual meetings of team representatives and representatives of the customer took place. As a result of working together with customer representatives all four development teams offered their SRS.

The sets R1-R4 of SRS indicators were formed for the each of four SRS and submitted for processing to the ANN. The results of ANN (predicted relative values of the characteristics), the calculated by MESSPI integrative indicators and degree of success of these projects implementation are in Table 1.

Table 1. Predicted relative values of characteristics, calculated integrative indicators and degree of success of four software projects implementation

Characteristics and indicators of software project	Values for Project1	Values for Project2	Values for Project3	Values for Project4
Cost Cs_{ANN}	0.8	0.22	0.39	0.59
Duration Dsp_{ANN}	0.9	0.19	0.41	0.57
Complexity Cx_{ANN}	0.75	0.31	0.37	0.62
Usability Ub_{ANN}	0.85	0.15	0.5	0.56
Cross-platformCp_{ANN}	0.87	0.21	0.47	0.57
Quality Qs_{ANN}	0.89	0.17	0.49	0.61
Integrative indicator Iip_{Sp}	*1,847*	*0,113*	*0,501*	*0,894*
The degree of success P_{Iip}	*0.7111*	*0,0435*	*0.1929*	*0.3442*

Thus, the results of Table 1 demonstrate that Project1 has the greatest predicted degree of success of implementation (71%) and Project2 has the smallest predicted degree of success of implementation (about 4%). Therefore the Project1 (SRS of Project1) was proposed to the developer and the customer for solution of their task.

If we will not take into account the compensation of low values of some characteristics by high values of other characteristics in the calculation of integrative indicator of the project, there is a risk for the obtaining of following results. Let the ANN gived certain values of characteristics for five different software projects. We show these values and the corresponding values of integrative indicators in Table 2.

The data of Table 2 show that all five software projects have the same integrative indicator Iip_{Sp}=0.894, but have significantly different relative values of characteristics. We need to check the convexity of the hexagons for all examined software projects for determination of value of indicator Ace_{Sp} - Table 3.

Table 2. Examples of compensation of characteristics for different software projects

Characteristics and indicators of project	Values for Pr.4	Values for Pr.5	Values for Pr.6	Values for Pr.7	Values for Pr.8
Cost Cs_{ANN}	0.59	0.7	1	1	0.93
Duration Dsp_{ANN}	0.57	0.57	0.57	0.57	0.57
Complexity Cx_{ANN}	0.62	0.62	0.62	0.62	0.62
Usability Ub_{ANN}	0.56	0.56	0.56	0.403	0.56
Cross-platformCp_{ANN}	0.57	0.57	0.57	0.57	0.57
Quality Qs_{ANN}	0.61	0.503	0.289	0.403	0.33
Integrative indicator Iip_{Sp}	*0.894*	*0.894*	*0.894*	*0.894*	*0.894*

Table 3. Testing of the stability and acceptability of compensatory effects of the characteristics for eight software projects

Values	Pr.1	Pr.2	Pr.3	Pr.4	Pr.5	Pr.6	Pr.7	Pr.8
Sine of angle Qs	+	+	+	+	+	-	+	-
Sine of angle Cs	+	+	+	+	+	+	+	+
Sine of angle Cx	+	+	+	+	+	+	+	+
Sine of angle Dsp	+	+	+	+	+	+	+	+
Sine of angle Ub	+	+	+	+	+	+	+	+
Sine of angle Cp	+	+	+	+	+	+	+	+
Indicator Ace_{Sp}	True	True	True	True	True	*False*	True	*False*

The testing of the stability and acceptability of compensations of characteristics of software projects showed that for Project6 and Project8 the characteristics are unstable, i.e. compensations of these characteristics are unacceptable.

4 Conclusions

This paper shows: the need of deepening of the SRS analysis; the dependence of quality and success of software project implementation on the SRS; the actuality and importance of the skill of evaluation of software project implementation success based on the SRS; the need of support of the choice of the best SRS for the project.

The authors first proposed the method of evaluating the success of software project implementation based on analysis of specification using neuronet information technologies. MESSPI differs from the known methods (analysed in [8-13]) that provides the prediction of the success of software projects implementation based on only SRS. The practical significance of the proposed method is the support in the comparison of software projects on the basis of SRS, the choice of the best SRS of

project, and control for SRS quality also (SRS quality is very importance, as known [14]). The proposed method is suitable only for software projects, for which SRS are existing and available. This method helps to "cut off" the software projects with failed SRS, because, as shown above, the software projects with failed requirements and specifications can not be successfull at the implementation.

The authors have following perspectives for future researches: 1) increasing of the veracity of ANN functioning for increasing of the MESSPI veracity; 2) selection of variant component for ANN; 3) providing recommendations about that is necessary to be changed in the SRS, that project became successful; 4) development of information technology for prediction of characteristics and evaluation of success of software project implementation based on the SRS analysis; this information technology should support: the SRS indicators collection, the processing of this data by ANN, the collection of the relative values of characteristics, the calculation of the integrative indicator and the degree of success of the software project implementation, and testing of the stability and acceptability of characteristics compensations.

References

1. The Standish Group International: CHAOS Manifesto – Think big, act small. Technical report, CHAOS Knowledge Center (2013)
2. Bourque, P., Fairley, R.: Guide to the software engineering body of knowledge (SWEBOK): Version 3.0. A project of the IEEE Computer Society (2014)
3. McConnell, S.: Code complete. Microsoft Press (2013)
4. Pomorova, O., Hovorushchenko, T.: The modern problems of software quality evaluation. Radioeletronic and computer systems. 5, 319-327 (2013) [in Ukrainian]
5. Levenson, N.G.: Systemic factors in software-related spacecraft accidents. In: AIAA Space Conference and Exposition, pp.1-11 (2001)
6. Levenson, N.G.: Software challenges in achieving space safety. Journal of the British Interplanetary Society. 62, 265-272 (2009)
7. Ishimatsu, T., Levenson, N., Thomas, J., Fleming, C., Katahira, M., Miyamoto, Y., Ujiie, R.: Hazard analysis of complex spacecraft using systems-theoretic process analysis. Journal of Spacecraft and Rockets. 51, 509-522 (2014)
8. Maedche, A., Botzenhardt, A., Neer, L.: Software for people: fundamentals, trends and best practices. Springer-Verlag Berlin Heidelberg, Berlin (2012)
9. Krasiy, A.: Modelling of process of prediction of software characteristics based on the analysis of specifications. Computer-Integrated Technologies: Education, Science, Industry. 66-76 (2014) [in Ukrainian]
10. Fenton, N.: Software metrics: A rigorous approach (3rd edition). CRC Press (2014)
11. Chen, A., Beatty, J.: Visual models for software requirements. MS Press, Washington (2012)
12. Fatwanto, A.: Software requirements specification analysis using natural language processing technique In: International Conference on Quality in Research, pp.105-110 (2013)
13. Rehman, T., Khan, M.N.A., Riaz, N.: Analysis of requirement engineering processes, tools/techniques and methodologies. I.J. Information Technology and Computer Science. 40-48 (2013)
14. IEEE 830-1998. Recommended practice for software requirements specifications (1998)

Calculation Method for a Computer's Diagnostics of Cardiovascular Diseases Based on Canonical Decompositions of Random Sequences

Igor P. Atamanyuk[1], Yuriy P. Kondratenko[2]

[1] Mykolaiv National Agrarian University, Commune of Paris str. 9,
54010 Mykolaiv, Ukraine
atamanyukip@mnau.edu.ua
[2] Petro Mohyla Black Sea State University, 68th Desantnykiv Str. 10,
54003 Mykolaiv, Ukraine
yuriy.kondratenko@chdu.edu.ua

Abstract. The canonical decomposition of sequence describing the change of cardiograms is put in the basis of the method for a computer system of disease diagnostics. Obtained criterion of the solution of the problem of electrocardiograms classification is considerably simpler than the known criterion of making decision on the basis of the criterion of the maximum of density of distribution. The transition from multi-dimension density distribution to producing of unidimensional densities that allows to use random number of parameters of electrocardiograms for diagnostics is offered to carry out. The results of numerical experiment confirm the effectiveness of the offered method and high reliability of the processes of identification of cardiovascular diseases identification on the basis of its usage.

Keywords: calculation method, medical diagnostics, electrocardiogram, random sequence, canonical decomposition.

Key Terms: computation, mathematical model.

1 Introduction

At present, cardiovascular diseases head the list among the most widespread and dangerous diseases of modernity [1]. According to the data of the world Health Organization the death rate because of heart diseases in Ukraine reaches 64%, in the USA heart disease affects more than 800 000 people annually. At present the number of heart diseases among capable of working population sharply increased (quite often the age of the sick person with cardiac infarction doesn't exceed 23-25 years).

As heart diseases belong to the diseases which course and results of treatment directly depend on timely detection and elimination of pathological deviations the reliable diagnostics is the most important and primary task in the problem of cardiovascular diseases. As of today a great number of approaches [2-12] for the solving of the

given task with the usage of different mathematical methods including statistical methods, methods of computational intelligence, fuzzy logic, neural network modeling algorithms and others are worked out.

Let us consider some related works concerning the methods for analysis of electrocardiograms using automated techniques, modern information technologies and computer systems. For example, such investigations were started at the University of Glasgow (Uni-G), United Kingdom more than 40 years ago and are continuing as Uni-G ECG Analysis Program [13] based on development of different approaches, in particular: methods for processing waveforms recorded in groups of three leads simultaneously, 12-lead ECG analysis program, optional approaches to computing the average QRS cycle including a simple mean, a weighted mean and a median beat, rhythm analysis, Brugada pattern, neural networks, rule based criteria, software diagnostic criteria based on age, sex, race, clinical classification, drug therapy and so on.

A dynamic hybrid architecture is descripted in [14] for ECG data analysis, combining the fuzzy with the connectionist approach. The data abstraction is performed by a layer of Radial Basis Function (RBF) units and the upcoming classification is carried out by a classical two-layer feedforward neural network. For the evaluation a large clinically validated ECG database is explored, but a more detailed description of the input space using a larger number of RBF units does not grant sufficient improvements.

Leiden ECG Analysis and Decomposition Software (LEADS) was developed [15] at the Leiden University Medical Center, The Netherlands as a MATLAB program for research oriented ECG/VCG analysis. LEADS focuses on the determination of a low-noise representative averaged beat (QRST complex), in which multiple parameters can be measured, paying special attention to the T wave. LEADS generates a default selection of beats for subsequent averaging.

The paper [16] presents the current status of principal component analysis (PCA) for ECG signal processing and describes the relationship between PCA and Karhunen-Loeve transform.

Several ECG applications based on PCA techniques have been successfully employed, including data compression, ST-T segment analysis for the detection of myocardial ischemia and abnormalities in ventricular repolarization, extraction of atrial fibrillatory waves for detailed characterization of atrial fibrillation, and analysis of body surface potential maps.

Advances in sensor technology, personal mobile devices, wireless broadband communications, and Cloud computing are enabling real-time collection and dissemination of personal health data to patients and health-care professionals anytime. This approach was proposed in [17] for creating an autonomic cloud environment for hosting ECG data analysis services.

A solution in [18] leverages the advance in multi-processor system-on-chip architectures, and is centered on the parallelization of the ECG computation kernel.

The article [19] reviewed time domain, frequency domain, premature complexes detection, heart rate variability, and nonlinear ECG analysis based methods.

Several different approaches for ECG analysis are based on a chaos theory [20], a combination of statistical, geometric, and nonlinear heart rate variability features [21],

a semantic web ontology and heart failure expert system [22], learning system based on support vector machines [23], signal averaging method, multivariate analysis [24], RPCA - recursive principal component analysis [25], nonlinear PCA neural networks [26], cluster analysis, SPSA - simultaneous perturbation stochastic approximation method [27], ABT - Amplitude Based Technique, FDBT - First Derivative Based Technique, SDBT - Second Derivative Based Technique [28], Hilbert transform [29] and so on.

At the same time each from above-mentioned methods has its disadvantages and limitations. Just therefore the necessity of the working out of new effective methods of medical diagnostics didn't lose its actuality.

2 Statement of the problem

One of the most widespread methods of diagnostics and detection of cardiovascular diseases is an electrocardiography, a method of graphic registration of the characteristics of the electric field of a heart and their changes in the process of heart contractions. Electrocardiogram is characterized with a set of teeth by time and amplitude parameters of which the diagnosis is done. Taking into account that changing of the parameters of electrocardiogram has accidental character the problem of the classification of the realization of random sequence (some disease or absence of a disease correspond to every class) is the mathematical content of heart diseases diagnostics. For the purpose of the increase of the reliability of the diagnostics of cardiovascular diseases it is necessary to work out on the basis of the theory of random sequences the method of electrocardiogram recognition with taking complete account of their stochastic qualities.

3 Solution

The object of investigation is the random consequence $\{X\} = \{X(1), X(2), ..., X(12)\}$ with twelve elements each of which corresponds to some the most informative parameter of the electrocardiogram Fig. 1 (as appropriate the number of parameters can be increased): $X(1)$ is the width of the tooth P; $X(2)$ is the height of the tooth P; $X(3)$ is the interval P–Q; $X(4)$ is the height of the tooth Q; $X(5)$ is the interval QRS; $X(6)$ is the height of the first tooth R; $X(7)$ is the height of the second tooth R; $X(8)$ is the height of the tooth S; $X(9)$ is the interval Q-T; $X(10)$ is the height of the tooth T; $X(11)$ is the duration of the first cycle of the cardiogram; $X(12)$ is the duration of the second cycle of the cardiogram.

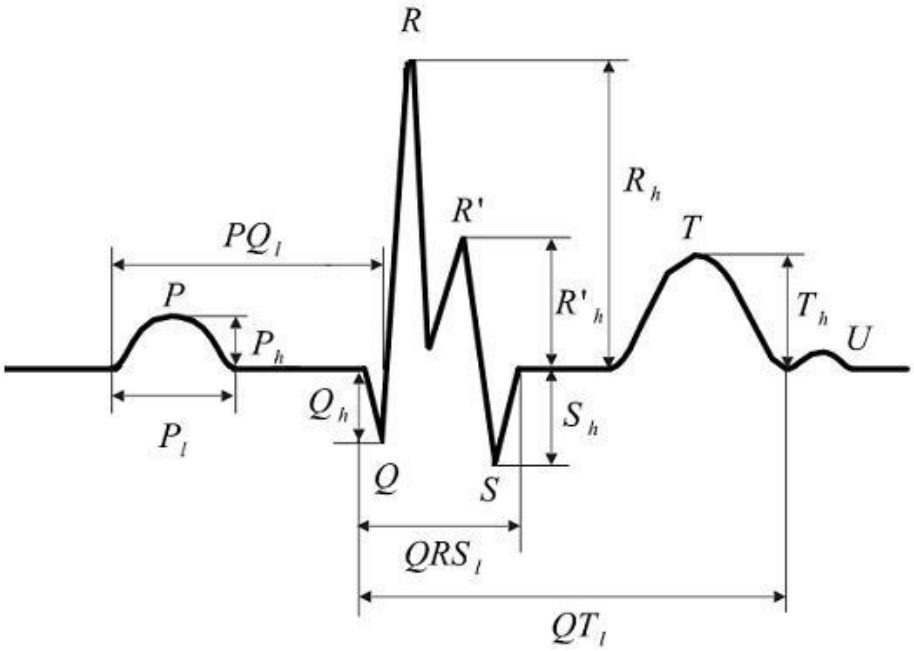


Fig. 1. Teeth and intervals on the cardiogram

As the result of electrocardiography conducting some sequence of values $x(i)$, $i=\overline{1,12}$ about which it is known a priori that it is generated by one of the random sequences $X^{(j)}(i)$, $i=\overline{1,12}, j=\overline{1,J}$ ($J-1$ of diseases and normal state) is obtained. It is necessary to define to which of these sequences exactly (to which of J classes) relates to given realization. Formulated in such a way the problem of recognition completely comes to standard Bayes approach but during the usage of Bayes criterion improbable (and that is why especially dangerous) diseases can not be recognized. Thereupon for solving of the problem of medical diagnostics the most acceptable is the criterion of the maximum of probability according to which during the observation of the realization $\vec{x}=\{x(1),x(2),...,x(12)\}$ that hypothesis is taken which meets the condition:

$$j^{*}=\arg\max_{j}\left\{f_{12}\left(\vec{x}/j\right)\right\}, \tag{1}$$

where $f_{12}\left(\vec{x}/j\right)$, $j=\overline{1,J}$ is the relative density distribution of the symptoms $\dot{x}$ provided that the realization belongs to the given class.

The problem of the recognition of random sequence realization comes to the determination of the belonging of the realization $\dot{x}$ to one of J given distributions $f_{12}\left(\vec{x}/j\right)$, $j=\overline{1,J}$.

Thus the following stage is the assessment of the unknown densities $f_{12}(\vec{x}/j),\ j=\overline{1,J}$ that in its turn taking into account the great number of the results of $x(i),\ i=\overline{1,12}$ observations is quite difficult and laborious procedure. Given problem in the context of linear relations is essentially simplified [30] during the transition from sequence $x(i),\ i=\overline{1,12}$ to the analysis of the set of uncorrelated values $v_i,\ i=\overline{1,I}$, which are determined from the canonical model of random sequence [31] presentation:

$$X(i)=\sum_{\nu=1}^{i} V_\nu \varphi_\nu(i),\ i=\overline{1,12}, \tag{2}$$

$$V_i = X(i)-\sum_{\nu=1}^{i-1} V_\nu \varphi_\nu(i),\ i=\overline{1,12}, \tag{3}$$

$$\varphi_\nu(i)=\frac{1}{D_\nu}\left\{M\left[X(\nu)X(i)\right]-\sum_{j=1}^{\nu-1} D_j \varphi_j(\nu)\varphi_j(i)\right\},\ \nu=\overline{1,I},\ i=\overline{\nu,I}. \tag{4}$$

$$D_i = M\left[X^2(i)\right]-\sum_{\nu=1}^{i-1} D_\nu \varphi_\nu^2(i),\ i=\overline{1,12}\ , \tag{5}$$

where $\varphi_\nu(i),\ \nu,i=\overline{1,I}$ is nonrandom coordinate function: $\varphi_\nu(\nu)=1,\ \varphi_\nu(i)=0$, if $\nu>i$.

In this case the substitution of $\vec{x}$ for vector $\dot{v}$ taking into account $f_I(\vec{v}/j)=\prod_{i=1}^{12} f_1(v_i/j),\ j=\overline{1,J}$ allows to put down the criterion of decision making in the following form:

$$j^*=\arg\max_j\left\{\prod_{i=1}^{12} f_1(v_i/j),\ j=\overline{1,J}\right\}. \tag{6}$$

The problem of recognition thus comes to consecutive approximation of twelve one-dimensional densities of distribution. The stochastic algorithm of diagnostics becomes simpler essentially but the transition from the vector $\dot{x}$ to the vector $\dot{v}$ is possible provided that the random sequences $\left\{X(i)/j\right\},\ i=\overline{1,12},\ j=\overline{1,J}$ have only linear relations. Taking down of the limitations of the random sequences $X^{(j)}(i),\ i=\overline{1,12},\ j=\overline{1,J}$ normal distribution is possible as a result of the usage of the corresponding nonlinear canonical decomposition [32-35]:

$$V_i^{(\lambda)}=X^{\lambda}(i)-\sum_{\nu=1}^{i-1}\sum_{j=1}^{N} V_\nu^{(j)}\beta_{\lambda\nu}^{(j)}(i)-\sum_{j=1}^{\lambda-1} V_i^{(j)}\beta_{\lambda i}^{(j)}(i),\ i=\overline{1,12}; \tag{7}$$

$$D_\lambda(i) = M\left[X^{2\lambda}(i)\right] - \sum_{\mu=1}^{i-1}\sum_{j=1}^{N} D_j(\mu)\left\{\beta_{\lambda\mu}^{(j)}(i)\right\}^2 - \sum_{j=1}^{\lambda-1} D_j(i)\left\{\beta_{\lambda i}^{(j)}(i)\right\}^2,\ i=\overline{1,12}; \quad (8)$$

$$\beta_{h\nu}^{(\lambda)}(i) = \frac{1}{D_\lambda(\nu)}\left(M\left[X^{\lambda}(\nu)X^{h}(i)\right] - \sum_{\mu=1}^{\nu-1}\sum_{j=1}^{N} D_j(\mu)\beta_{\lambda\mu}^{(j)}(\nu)\beta_{h\mu}^{(j)}(i) - \right.$$
$$\left. - \sum_{j=1}^{\lambda-1} D_j(\nu)\beta_{\lambda\nu}^{(j)}(\nu)\beta_{h\nu}^{(j)}(i)\right),\ \lambda=\overline{1,N}, \nu=\overline{1,i}. \quad (9)$$

Taking into account different qualities of random sequences $\left\{X(i)/j\right\}$, $i=\overline{1,12}, j=\overline{1,J}$ parameters of the canonical decomposition (7)-(9) are unique for each of the investigated sequences. The advantage of the decomposition (7)-(9) usage is that their independence follows from noncorrelatedness $V_i^{(N)}$, $i=\overline{1,I}$ as all stochastic relations of much lower order are removed from the given coefficients. Thus the same as in the previous case the conversion of the problem of recognition from twelve measured space of the characteristics $\left\{X(1),...,X(12)\right\}$ into the space of the characteristics $\left\{V_1^{(N)},...,V_{12}^{(N)}\right\}$ of the same dimension simplifies the procedure of the assessment of the densities of distribution $f_{12}\left(v_1^{(N)},...,v_I^{(N)}/j\right) = \prod_{i=1}^{12} f_1\left(v_i^{(N)}/j\right),\ j=\overline{1,J}$ that comes to the approximation of twelve unidimensional densities of distribution. The criterion of making decision takes the following form

$$j^* = \arg\max_j\left\{\prod_{i=1}^{12} f_1\left(v_i^{(N)}/j\right), j=\overline{1,J}\right\}. \quad (10)$$

The absence of the assumptions about the kind of the density distribution of the random values $\left\{V_1^{(N)},...,V_{12}^{(N)}\right\}$ comes to the necessity of the usage of nonparametric methods for their description. The simplest and the most effective approach under given conditions is the usage of nonparametric assessments of Parzen-type [36]:

$$f_L\left(v_i^{(N)}\right) = \frac{1}{dL}\sum_{l=1}^{L} g\left(u_l\right), \quad (11)$$

where $u_l = d^{-1}\left(v_i^{(N)} - v_{i,l}^{(N)}\right)$, $v_{i,l}^{(N)}$, $l=\overline{1,L}$ are the realizations of the random value $V_i^{(N)}$, $g\left(u_l\right)$ is a certain weigh function (kernel); d is a constant (coefficient of blurriness).

The choice in the capacity of the function of the kernel of $g(u)$ of steady density distribution allows to write down the expression for the assessment of the density distribution of $V_i^{(N)}$ in the following form:

$$f_L\left(v_i^{(N)}\right)=\frac{1}{dL}\sum_{l=1}^{L} g_l\left(v_i^{(N)}\right),$$

where

$$g_l\left(v_i^{(N)}\right)=\begin{cases}0{,}5, & v_{i,l}^{(N)}-d\le v_i^{(N)}\le v_{i,l}^{(N)}+d,\\ 0, & \left|v_i^{(N)}-v_{i,l}^{(N)}\right|>d,\end{cases} \quad l=\overline{1,L};$$

$$d=0{,}5\sup_l\left|v_{i,l}^{(N)}-v_{i,l-1}^{(N)}\right|,\ v_{i,l}^{(N)}>v_{i,l-1}^{(N)},\ l=\overline{2,L}.$$

The method of diagnostics of cardiovascular diseases on the basis of the offered algorithm and criterion of making decisions presupposes the fulfillment of the following phases:

Phase 1. Collection of statistic information about each investigated random sequence $X^{(j)}(i)$, i=$\overline{1,I}$, j=$\overline{1,J}$;

Phase 2. Calculation on the basis of the accumulated realizations $x_l^{(j)}(i)$, $i=\overline{1,I}$; $l=\overline{1,L_j}$; $j=\overline{1,J}$ for the investigated sequences $X^{(j)}(i)$, i=$\overline{1,I}$, j=$\overline{1,J}$ discretized moment functions $M\left[X_l^{\lambda}(\nu)X_h^{\mu}(i)\right]$;

Phase 3. Forming for each sequence $X^{(j)}(i)$, i=$\overline{1,I}$, j=$\overline{1,J}$ the canonical decomposition (7);

Phase 4. Obtaining on the basis of statistic information the assessments of one-dimensional densities of the distribution of the random coefficients of the canonical decompositions of the random sequences $X^{(j)}(i)$, i=$\overline{1,I}$, j=$\overline{1,J}$;

Phase 5. Decomposition of the recognizable realization by canonical expressions; calculation of the values of one-dimensional densities of distribution of coefficients formed as a result of decompositions; determination of the belonging of the realization of a certain random sequence $X^{(j^*)}(i)$, i=$\overline{1,I}$ (diagnostics of a disease) with the help of a rule (10);

Phase 6. Entry of the recognized realization $x^{(j^*)}(i)$, i=$\overline{1,I}$ into the base of statistical data of the corresponding random sequence $X^{(j^*)}(i)$, i=$\overline{1,I}$.

The scheme of the functioning of the system of cardiovascular diseases diagnostics is represented in Fig. 2.

In modern medicine more than one hundred different cardiovascular diseases are classified [1]. Developed six-stage algorithm is tested on five the most widespread diagnoses: "healthy heart" – is a random sequence $\{X(i)/1\}$, $i=\overline{1,12}$; "hypertrophy of myocardium" - $\{X(i)/2\}$, $i=\overline{1,12}$; "severe arrhythmia" - $\{X(i)/3\}$, $i=\overline{1,12}$; "stenocardia of the 2d functional class" - $\{X(i)/4\}$, $i=\overline{1,12}$; "neurocirculatory dystonia of light degree" - $\{X(i)/5\}$, $i=\overline{1,12}$.The check of the statistical hypothesis about the independence of random coefficients of the canonical decomposition (7) on the basis of the criterion χ^2 showed the validity of the hypothesis by $N=3$ for all three sequences with the probability not less than $P_D=0,98$. Thus the decomposition (7) with the corresponding set of coordinate functions $\beta_{h\nu}^{(\lambda)}(i)$, $h,\lambda=\overline{1,3}$, $\nu,i=\overline{1,12}$ modifies into the adequate model of the investigated random sequence $\{X(i)/j\}$, $i=\overline{1,12}$, $j=\overline{1,3}$. For example, in Table 1 values $\beta_{1\nu}^{(1)}(i)$, $\nu,i=\overline{1,12}$ for $\{X(i)/3\}$, $i=\overline{1,12}$ are represented.

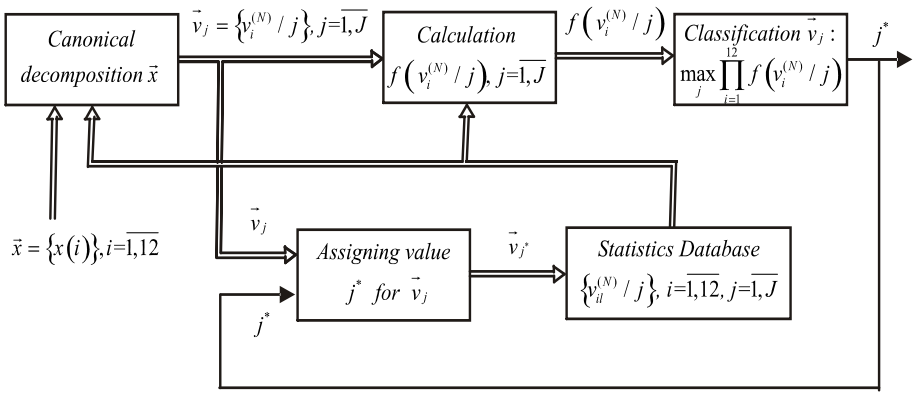


Fig. 2. Scheme of functioning of the computer system of cardiovascular diseases diagnostics

Recognition of the diagnoses was done on the basis of 200 different cardiograms for each disease. Comparative results of recognition of the diagnoses (a) on the basis of the developed by the authors calculating method, (b) on the basis of neuronic network [37] synthesized with the usage of Daubechies wavelet function of the 4th degree and Levenberg-Marquardt algorithm (for training) and (c) on the basis of the usage of fuzzy logic in medical diagnostics [3, 4] during the realization of the systems of fuzzy logic inference of Mamdani-type are presented in Table 2.

Neuronic network that was used in calculating experiment (Table 2) has the following pecularities.

1. Expressions for the determination of approximation coefficients and detailing of discrete wavelet transform are of the form [37]:

$$W_{\varphi}(j_0,k)=\frac{1}{M}\sum_{x} f(x)\varphi_{j_0,k}(x),$$

$$W_{\psi}(j,k)=\frac{1}{M}\sum_{x} f(x)\psi_{j,k}(x),$$

where $\varphi_{j,k}(x)$, $\psi_{j,k}(x)$ is a family of basic functions.

Table 1. Values of the coordinate function $\beta_{1\nu}^{(1)}(i)$ for random sequence $\{X(i)/3\}$, $i=\overline{1,12}$

	2	3	4	5	6	7	8	9	10	11	12
1	0,14	1,46	0,12	0,92	1,49	0,06	0,22	3,06	-0,36	7,11	5,66
2	1	6,50	0,34	3,81	5,70	0,37	1,56	12,5	-2,41	28,72	22,1
3	0	1	0,08	0,63	1,01	0,04	0,15	2,13	-0,24	4,91	3,92
4	0	0	1	4,22	9,07	0,72	1,12	14,1	-0,18	33,40	25,3
5	0	0	0	1	1,22	0,20	0,29	3,02	-0,25	6,42	5,46
6	0	0	0	0	1	0,08	0,15	1,61	-0,14	3,79	2,79
7	0	0	0	0	0	1	0,24	2,70	-0,53	6,16	5,05
8	0	0	0	0	0	0	1	5,69	-0,50	11,17	9,52
9	0	0	0	0	0	0	0	1	-0,07	2,12	1,82
10	0	0	0	0	0	0	0	0	1	-6,08	-4,1
11	0	0	0	0	0	0	0	0	0	1	0,83
12	0	0	0	0	0	0	0	0	0	0	1

2. Outcoming signal of each of separate neuron of outcoming layer was forming as

$$y(k)=\frac{1}{M} f\left(\sum_{i=0}^{K} w_{ki} f\left(\sum_{j=0}^{N} w_{ij} x\right)\right).$$

3. As activation function of each separate neuron continuous sigmoid bipolar function $f(x)=th(x)$ was being used.

In calculating experiment of the diagnostics of cardiovascular diseases on the basis of the realization of the mechanism of fuzzy logic inference [3,4] the following input parameters were used: x_1 - age of the sick; x_2 - double product of pulse on arterial

tension; x_3 - tolerance to physical activity; x_4 - increase of double product per one kilogram of the body weight of the sick; x_5 - increase of double product per one kilogram of physical exertion; x_6 - adenosinetriphosphoric acid; x_7 - adenosine diphosphoric acid; x_8 - adenylic acid; x_9 - coefficient of phosphorilation; x_{10} - maximal consumption of oxygen per one kilogram of the body weight of the sick; x_{11} - increase of double product in the response for submaximal physical exertion; x_{12} - coefficient of the ratio of lactic and pyruvic acid content.

Expressions for the determination of the diagnosis are of the form:

$$d = f_d(x_1, y, z),$$

$$y = f_y(x_2, x_3, x_4, x_5, x_{10}, x_{11}),$$

$$z = f_y(x_6, x_7, x_8, x_9, x_{12}),$$

where values d (diagnosis), y, z are determined with the help of the knowledge base mentioned in the works of professor A. P. Rotstein [3,4].

Table 2. Results of the diagnostics of cardiovascular diseases (% of correct solutions)

	Healthy heart	Hypertrophy of myocardium	Severe arrhythmia	Stenocardia of the 2d functional class	Neurocirculatory dystonia of light degree
Method on the basis of canonical expansion	100%	100%	100%	98%	97%
Method on the basis of neural network	89%	92%	94%	86%	83%
Method on the basis of fuzzy logic	91%	90%	93%	91%	89%

The results of numerical experiment confirm high effectiveness of the developed calculating method in the comparison to the methods of artificial intelligence at the expense of the usage of optimal parameters during the formation of the criterion of making decision.

The choice of Daubechies function of the 4th degree from the existing limited set of wavelet functions in the capacity of the parameter of neural network is not optimal for solving of the problem of cardiovascular diseases diagnostics (usage of other functions leads to the worsening of quality of problem solution [37]).

The results of the experiment on the basis of A. P. Rotstein's approach [3, 4] indicate that the absence of strict mathematical apparatus of fuzzy equation analysis doesn't allow to form optimal structure of fuzzy rules that naturally restricts the accuracy of cardiovascular diseases classification.

On the whole the basis of statistic data can be expanded by the way of the introduction of cardiogram information about wider class or about all existing types of cardiovascular diseases. This will allow to form on the basis of developed calculating method highly efficient information systems of cardiovascular diseases diagnostics for their actual usage in medical cardiologic centers, clinics and diagnostic establishments.

4 Conclusions

Therefore in the work the calculation method for a computer system of cardiovascular diseases diagnostics on the basis of the canonical decomposition of the random sequence of electrocardiogram change is offered. The use of the mechanism of canonical decompositions allowed to formulate the decisive rule of the maximum of the combined density distribution in the form of the production of one-dimensional densities of distribution that gives the possibility to use for diagnostics random quantity of electrocardiogram parameters. Besides canonical decomposition doesn't impose any essential limitations (linearity, stationarity, Markovian property etc.) on the class of investigated random sequences. Thereby the offered approach to the solution of the problem of cardiovascular diseases diagnostics allows to take into account the maximum stochastic characteristics of the electrocardiograms belonging to different cardiovascular diseases. The given results of modeling show the high reliability of cardiovascular diseases diagnostics on the basis of the offered method.

5 References

1. Organov R.G., Komarov Y.M., Maslennikova G.Y.: Demographic Problems as a Mirror of Nation's Health. J. Prophylactic Medicine 2, 3-8 (2009)
2. Kotov, Y.B.: New Mathematical Approaches to the Problems of Medical Diagnostics. Editorial EPCC, Moscow (2004)
3. Rotshtein, A.P.: Intellectual Technologies of Identification: Fuzzy Logic, Genetic Algorithms, Neuron Networks, UNIVERSUM, Vinnitsa (1999)
4. Rotshtein, A.P.: Medical Diagnostics on the Fuzzy Logic. Kontingent-Prim, Vinnitsa (1996)
5. Boyko V.V., Bodyansky E.V., Vinokurova E.A., Sushkov S.V., Pavlov A.A. Analysis of clinical data in medical research based on methods of computational intelligence. TO Exclusive, Kharkov (2008)
6. Abdel–Badeeh M. Salem, Mohamed Roushdy, Rania A.: A Case Based Expert System for Supporting Diagnosis of Heart Diseases. J. ICGST International Journal on Artificial Intelligence and Machine Learning 33–39 (2005)
7. Yezhov A., Chechetkin V.: Neural Networks in Medicine. J. Open Systems. 4, 34 – 37 (1997)

8. Dasilva P., Fortier P., Sethares K.: Electrocardiogram Classification Sensor System Supporting an Autonomous Mobile Cardiovascular Disease Detection Aid J. Sensors & Transducers 184, 92-100 (2015)
9. Niknazar M., Vahdat B.V., Mousavi S. R.: Detection of Characteristic Points of ECG using Quadratic Spline Wavelet Transform. Proceedings of the 3rd nternational Conference on Signals, Circuits and Systems (SCS'09), Medenine, Tunisia, 6-7 November, 1-6 (2009)
10. Sasikala P., Wahida Banu R.: Extraction of P wave and T wave in Electrocardiogram using Wavelet Transform. J. International Journal of Computer Science and Information Technologies 2, 489-493 (2011)
11. Ranjith P., Baby P., Joseph P.: ECG Analysis Using Wavelet Transform: Application to Myocardial Ischemia Detection. J. ITBM-RBM. 24, 44-47 (2011)
12. Lasted L.: Introduction into the Problem of Taking Decisions in Medicine. Mir, Moscow (1971)
13. Macfarlane, P. W., Devine, B., Clark, E.: The university of Glasgow (Uni-G) ECG analysis program. In Computers in Cardiology, IEEE, 451-454 (2005)
14. Silipo, R., Bortolan, G., Marchesi, C.: Design of hybrid architectures based on neural classifier and RBF pre-processing for ECG analysis. International Journal of Approximate Reasoning, 21(2), 177-196 (1999)
15. Draisma, H. H. M., Swenne, C. A., Van de Vooren, H., Maan, A. C., Hooft van Huysduynen, B., Van der Wall, E. E., Schalij, M. J. LEADS: an interactive research oriented ECG/VCG analysis system. In Computers in Cardiology, IEEE, 515-518 (2005)
16. Castells, F., Laguna, P., Sörnmo, L., Bollmann, A., Roig, J. M.: Principal component analysis in ECG signal processing. EURASIP Journal on Applied Signal Processing, 2007(1), 98-98 (2007)
17. Pandey, S., Voorsluys, W., Niu, S., Khandoker, A., Buyya, R.: An autonomic cloud environment for hosting ECG data analysis services. Future Generation Computer Systems, 28(1), 147-154 (2012)
18. Al Khatib, I., Bertozzi, D., Poletti, F., Benini, L., Jantsch, A., Bechara, M., ... Jonsson, S.: MPSoC ECG biochip: a multiprocessor system-on-chip for real-time human heart monitoring and analysis. In Proceedings of the 3rd Conference on Computing Frontiers, ACM, 21-28 (2006)
19. Poli, S., Barbaro, V., Bartolini, P., Calcagnini, G., Censi, F.: Prediction of atrial fibrillation from surface ECG: review of methods and algorithms. Annali dell'Istituto superiore di sanità, 39(2), 195-203 (2002)
20. Jovic, A., Bogunovic, N.: Feature extraction for ECG time-series mining based on chaos theory. In Information Technology Interfaces, 2007. ITI 2007. 29th International Conference on., IEEE, 63-68 (2007)
21. Jovic, A., Bogunovic, N.: Electrocardiogram analysis using a combination of statistical, geometric, and nonlinear heart rate variability features. Artificial intelligence in medicine, 51(3), 175-186 (2011)
22. Prcela, M., Gamberger, D., Jovic, A.: Semantic web ontology utilization for heart failure expert system design. Studies in health technology and informatics, (136), 851-6 (2008)
23. Jankowski, S., Oreziak, A.: Learning system for computer-aided ECG analysis based on support vector machines. International Journal of Bioelectromagnetism. ISBEM (2003).
24. Biel, L., Pettersson, O., Philipson, L., Wide, P.: ECG analysis: a new approach in human identification. Instrumentation and Measurement, IEEE Transactions on, 50(3), 808-812 (2001)
25. Pawar, T., Anantakrishnan, N. S., Chaudhuri, S., Duttagupta, S. P.: Impact analysis of body movement in ambulatory ECG. In Engineering in Medicine and Biology Society,

2007. EMBS 2007. 29th Annual International Conference of the IEEE, IEEE, 5453-5456 (2007)
26. Stamkopoulos, T., Diamantaras, K., Maglaveras, N., Strintzis, M. ECG analysis using non-linear PCA neural networks for ischemia detection. Signal Processing, IEEE Transactions on, 46(11), 3058-3067 (1998)
27. Gerencsér, L., Kozmann, G., Vágó, Z., Haraszti, K.: The use of the SPSA method in ECG analysis. Biomedical Engineering, IEEE Transactions on, 49(10), 1094-1101 (2002)
28. Fang, Q., Sufi, F., Cosic, I.: A mobile device based ECG analysis system. NTECH Open Access Publisher (2008)
29. Benitez, D., Gaydecki, P. A., Zaidi, A., Fitzpatrick, A. P.: The use of the Hilbert transform in ECG signal analysis. Computers in biology and medicine, 31(5), 399-406 (2001)
30. Kudritsky V.D.: Filtering, extrapolation and recognition realizations of random functions. FADA Ltd., Kyiv (2001)
31. Pugachev V. S.: The Theory of Random Functions and its Application. Fitmatgiz, Moscow (1962)
32. Atamanyuk, I.P., Kondratenko Y. P.: The Algorythm of Optimal Nonlinear Extrapolation of the Realizations of Random Process with the Filtration of Errors Changes. J. Electronic Modelling 4, 23-40 (2012)
33. Atamanyuk, I.P., Kondratenko, V.Y., Kozlov, O.V., Kondratenko, Y.P.: The algorithm of optimal polynomial extrapolation of random processes, Modeling and Simulation in Engineering, Economics and Management, LNBIP 115, Springer, New–York, 78-87 (2012)
34. Atamanyuk I. P.: The Algorithm to Determine the Optimal Parameters of a Wiener Filter-extrapolator for Non-stationary Stochastic Processes Observed with Errors. J. Cybernetics and Systems Analysis 4, 154-159 (2011)
35. Atamanyuk I.P., Kondratenko Y.P.: The Synthesis of Optimal Linear Stochastic Systems of Control on the Basis of the Apparatus of Canonical Decompositions of Random Sequences. J. Controlling Systems and Machines. 1, 8–12 (2012)
36. Parzen, E.: On the estimation of probability density function and the mode. J. Analysis of Mathematical Statistics 33, 1065-1076 (1962)
37. Grigoriev D.S, Spitsin V.G. The application of neural network and discrete wavelet transform for the analysis and classification of electrocardiograms. J. Bulletin of the Tomsk Polytechnic University 5, 57-61 (2012)

Synthesis of Time Series Forecasting Scheme Based on Forecasting Models System

Fedir Geche[1], Vladyslav Kotsovsky[2], Anatoliy Batyuk[3], Sandra Geche[4], and Mykhaylo Vashkeba[1]

[1] Uzhhorod National University, Department of Cybernetics and Applied Mathematics, Uzhhorod, Ukraine
(fgeche@hotmail.com, vashkebam1991@gmail.com)
[2] Uzhhorod National University, Department of Information Management Systems, Uzhhorod, Ukraine
kotsavlad@gmail.com
[3] Lviv Polytechnic National University, Department of Automatic Control Systems, Lviv, Ukraine
abatyuk@gmail.com
[4] Uzhhorod National University, Department of Economic Theory, Uzhhorod, Ukraine
sandra.geche@gmail.com

Abstract. This article is dedicated to the development of time series forecasting scheme. It is created based on the forecasting models system that determines the trend of time series and its internal rules. The developed scheme is synthesized with the help of basic forecasting models "competition" on a certain time interval. As a result of this "competition", for each basic predictive model there is determined the corresponding weighting coefficient, with which it is included in the forecasting scheme. Created forecasting scheme allows simple implementation in neural basis. The developed flexible scheme of forecasting of economic, social, environmental, engineering and technological parameters can be successfully used in the development of substantiated strategic plans and decisions in the corresponding areas of human activity.

Keywords. Trend, forecasting model, time series, functional, step of forecast, autoregression, neural element, neural network.

Key Terms. MachineIntelligence, DecisionSupport, MathematicalModel

1 Introduction

At the present stage, for effective management of enterprises it is necessary to be able to predict the major trends in social and economic systems, the main economic indicators characterizing financial position and efficiency of the use of companies' production resources.

Estimates and forecasts of the financial condition of the company make it possible to find additional resources, to increase its profitability and solvency.

Problems of the analysis and the forecast of financial condition of the company by means of corresponding indicators are an actual task, because on the one hand this is the result of the company, on the other it defines the preconditions for the development of the company. Qualitative forecast gives us an opportunity to develop reasonable strategic plans for economic activity of enterprises.

Under market conditions, the adequate forecasting and capacity planning of enterprises are impossible without working out economic and mathematical models that describe the use of available resources during the operation of enterprises.

To determine strategies for enterprise development, calculation of forecasts of economic indicators and factors of organizations plays an important role. If there is reliable information about the company in the past, mathematical methods can be applied to obtain necessary forecasts. These methods depend on the objectives and detailed forecast factors; they also depend on the environment.

Various aspects of the theory, practice, and forecast of financial condition of a company have been the subject of research of many domestic and foreign scientists, such as Blank I.A [1], Heyets V.M. [2], Zaychenko Y.P. [3], Ivakhnenko V.M. [4], Ivakhnenko O.G. [5], Yarkina N.M. [6], Tymashova L. [7], Stepanenko O.P. [8], Tkachenko R.O. [9], Matviichuk A.V. [10], Hanke J.E. [11], Lewis C.D. [12], Box G.E. [13].

When forecasting the indicators by which the financial position or efficiency of the company's production resources use are determined, it is impossible to point out a single "the best" method of prediction because the internal laws (trends) of various indicator systems are different and there arises the problem of choosing the method of forecasting the studied indicator system.

Therefore, the development of new forecasting models of corresponding systems of indicators is an actual and important problem.

The aim of the study is to develop an efficient scheme of time series prediction that automatically (in the course of its training) adjusts to the appropriate system of economic, social, environmental, and engineering parameters, and it can be successfully used in the development of high-quality strategic plans in the branch of economy, environment, and for forecast of different natural processes.

The research methodology includes the method of least squares, exponential smoothing method, iterative techniques of minimization of functionals, and methods of synthesis of neural-network schemes.

2 Synthesis of Forecasting Schemes of Time Series

Let $v_1, v_2, ..., v_t, ..., v_n$ be a time series. Prognostic value $\tilde{v}_t$ of the element v_t at the instant of time t can be written as follows [14-16]

$$\tilde{v}_t = f(a_1, ..., a_r, v_{t-1}, ..., v_{t-k}, t), \tag{1}$$

where $a_1, ..., a_r$ are the model parameters, k is the depth of prehistory. To find the parameters $a_1, ..., a_r$, we constructed the functional

$$L(a_1, ..., a_r) = \sum_{t=1}^{n} \left(v_t - \tilde{v}_t\right)^2, \tag{2}$$

which is usually to be minimized. Let $a_1^*, ..., a_r^*$ are the values of parameters $a_1, ..., a_r$ for which the functional L takes its minimum value. Then the prognostic value $\tilde{v}_{n+\tau}$ of the model f with optimal parameters $a_1^*, ..., a_r^*$ is determined as follows

$$\tilde{v}_{n+\tau} = f(a_1^*, ..., a_r^*, v_{t-1}, ..., v_{t-k}, n+\tau), \tag{3}$$

where τ is the step of the forecast. Depending on the type of the function f with the parameters $a_1^*, ..., a_r^*$, we have different optimal forecasting models of time series.

To build a predictive scheme, at the beginning let us consider the autoregression method by means of which we define the optimal step of the prehistory k_τ^* for the given time series v_t with the fixed step of the forecast τ. In the autoregression model, it is assumed that the indicator value v_t at the instant of time t depends on $v_{t-\tau}, v_{t-\tau-1}, ..., v_{t-\tau-k_\tau+1}$, where k_τ is the parameter of the prehistory with fixed τ. The prognostic value $\tilde{v}_{n+\tau}$ by the autoregression method is found according to the following model

$$\tilde{v}_{n+\tau} = a_1^{(\tau)} v_n + a_2^{(\tau)} v_{n-1} + ... + a_{k_\tau}^{(\tau)} v_{n-k_\tau+1}. \tag{4}$$

To determine the optimal values of the parameters $a_t^{*(\tau)} (t = 1,2,...,k_\tau)$ for a fixed τ $\left(= {}_0 \right)$, we minimize the functional

$$L(a_1^{(\tau)}, ..., a_{k_\tau}^{(\tau)}) = \sum_{t=k_\tau+\tau}^{n} \left(v_t - a_1^{(\tau)} v_{t-\tau} - ... - a_{k_\tau}^{(\tau)} v_{t-\tau-k_\tau+1}\right)^2, \tag{5}$$

i.e. we solve the system of equations

$$\frac{\partial L}{\partial a_i^{(\tau)}} = 0, t = 1,2,...,k_\tau. \quad (6)$$

Let $a_1^{*(\tau)},...,a_{k_\tau}^{*(\tau)}$ be a solution of the system (6). Then, according to (4) we have

$$\tilde{v}_t = a_1^{*(\tau)} v_{t-\tau} + a_2^{*(\tau)} v_{t-\tau-1} + ... + a_{k_\tau}^{*(\tau)} v_{t-\tau-k_\tau+1}, \quad (7)$$

where $t \geq k_\tau + \tau.$

It is obvious that the variable $\tilde{v}_t$ for a fixed value of τ $(\tau = \tau_0)$ depends on the parameter $k_\tau (1 \leq k_\tau \leq n - \tau)$. To determine the optimal value of the prehistory parameter k_τ for $\tau = \tau_0$ for the given time series v_t, let us consider the variables

$$\delta_1 = \frac{1}{n-\tau} \sum_{t=\tau+1}^{n} \left(v_t - a_1^{*(\tau)} v_{t-\tau} \right)^2,$$

$$\delta_2 = \frac{1}{n-\tau-1} \sum_{t=\tau+2}^{n} \left(v_t - a_1^{*(\tau)} v_{t-\tau} - a_2^{*(\tau)} v_{t-\tau-1} \right)^2,$$

.....

$$\delta_{n-\tau} = \left(v_n - a_1^{*(\tau)} v_{n-\tau} - ... - a_{n-\tau}^{*(\tau)} v_1 \right)^2$$

Thus we obtain $\min\{\delta_1, \delta_2, ..., \delta_{n-\tau}\} = \delta_{k_\tau^*}$. The variable k_τ^* determines the optimal value of the prehistory parameter in the autoregression model for a fixed τ $(\tau = \tau_0)$.

After determining the k_τ^* for a fixed $(\tau = \tau_0)$, consider the main base forecasting models $M_1, M_2, .. M_q$ of time series with the fixed step of the forecast τ, i.e. models on the bases of which a new forecasting scheme are synthesized. Using the results of the forecasting models mentioned above on the time interval $t = n - k_\tau^* + 1, n - k_\tau^* + 2, \ldots, n$, we draw the following table

Table 1. The Prognostic Values of Time Series

Forecasting Models	**Elements of Time Series** v_t			
	$v_{n-k_\tau^*+1}$	$v_{n-k_\tau^*+2}$	$\cdots$	v_n
M_1	$\tilde{v}^{(1)}_{n-k_\tau^*+1}$	$\tilde{v}^{(1)}_{n-k_\tau^*+2}$	$\cdots$	$\tilde{v}^{(1)}_n$
M_2	$\tilde{v}^{(2)}_{n-k_\tau^*+1}$	$\tilde{v}^{(2)}_{n-k_\tau^*+2}$	$\cdots$	$\tilde{v}^{(2)}_n$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
M_q	$\tilde{v}^{(q)}_{n-k_\tau^*+1}$	$\tilde{v}^{(q)}_{n-k_\tau^*+2}$	$\cdots$	$\tilde{v}^{(q)}_n$

In each column $v_{n-k_\tau^*+1}, v_{n-k_\tau^*+2},\ldots,v_n$ of Table 1, we can find the least squared difference of the prognostic and the actual values of the corresponding time series terms. Mathematically this can be written as following:

$$\text{let } j_1 = n - k_\tau^* + 1 \quad \text{and}$$

$$\varepsilon_1 = \min\left\{(v_{j_1} - \tilde{v}^{(1)}_{j_1})^2, (v_{j_1} - \tilde{v}^{(2)}_{j_1})^2, \ldots, (v_{j_1} - \tilde{v}^{(q)}_{j_1})^2\right\},$$

$$j_2 = n - k_\tau^* + 2 \quad \text{and}$$

$$\varepsilon_2 = \min\left\{(v_{j_2} - \tilde{v}^{(1)}_{j_2})^2, (v_{j_2} - \tilde{v}^{(2)}_{j_2})^2, \ldots, (v_{j_2} - \tilde{v}^{(q)}_{j_2})^2\right\},$$

……

$$j_{k_\tau^*} = n \text{ and}$$

$$\varepsilon_{k_\tau^*} = \min\left\{(v_n - \tilde{v}^{(1)}_n)^2, (v_n - \tilde{v}^{(2)}_n)^2, \ldots, (v_n - \tilde{v}^{(q)}_n)^2\right\}$$

Define the sets $I_1, I_2, \ldots, I_{k_\tau^*}$ as follows

$$I_1 = \left\{ i \in \{1,2,\ldots,q\} \middle| \varepsilon_1 = (v_{j_1} - v_{j_1}^{(i)})^2 \right\},$$

$$I_2 = \left\{ i \in \{1,2,\ldots,q\} \middle| \varepsilon_2 = (v_{j_2} - v_{j_2}^{(i)})^2 \right\},$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$I_{k_\tau^*} = \left\{ i \in \{1,2,\ldots,q\} \middle| \varepsilon_{k_\tau^*} = (v_n - v_n^{(i)})^2 \right\}$$

and draw the table

Table 2. Parameters for Determining the Weighting Coefficients of the Model

Forecasting Models	j_1	j_2	$\ldots$	$j_{k_\tau^*}$	**Resultant Column**
M_1	a_{11}	a_{12}	$\ldots$	$a_{1k_\tau^*}$	S_1
M_2	a_{21}	a_{22}	$\ldots$	$a_{2k_\tau^*}$	S_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
M_q	a_{q1}	a_{q2}	$\ldots$	$a_{qk_\tau^*}$	S_q

where

$$a_{ps} = \begin{cases} \beta^{k_\tau^* - s}, if\, s \in I_s, \\ 0, if\, s \notin I_s, \end{cases}$$

$$S_p = \sum_{j=1}^{k_\tau^*} a_{pj}, 0 < \beta \le 1, (p = 1,2,\ldots,q, s = 1,2,\ldots,k_\tau^*).$$

With the help of $S_p = S_p(\beta)$ and $S(\beta) = \sum_{p=1}^{q} S_p(\beta)$ we determine the weighting coefficients of the forecasting models $M_p (p \le q)$, with which these models are included in the following forecasting scheme

$$\tilde{v}_{n+\tau} = \frac{S_1(\beta)}{S(\beta)} \tilde{v}_{n+\tau}^{(1)} + \frac{S_2(\beta)}{S(\beta)} \tilde{v}_{n+\tau}^{(2)} + \ldots + \frac{S_q(\beta)}{S(\beta)} \tilde{v}_{n+\tau}^{(q)}. \quad (8)$$

The coefficients of the forecasting models in the scheme (8) depend on the parameter β that determines the influence of the element v_t upon the prognostic value $\tilde{v}_{n+\tau}$. The more remote element v_t is from the prognostic point $\tilde{v}_{n+\tau}$, the less is its influence on the prognostic value $(0<\beta<1)$. In the case of $\beta=1$, all points of time series v_t are equivalent, i.e. in the model (8) the distance of the element v_t from the prognostic point $\tilde{v}_{n+\tau}$ is not taken into account.

Synthesis of the predictive scheme (8) will be completed in the course of training its concerning β. For this purpose, we construct the functional

$$L(\beta)=\sum_{i=1}^{k_\tau^*}(v_{j_i}-\frac{S_1(\beta)}{S(\beta)}\tilde{v}_{j_i}^{(1)}-\ldots-\frac{S_{q+r}(\beta)}{S(\beta)}\tilde{v}_{j_i}^{(q)})^2,\ (j_i=n-k_\tau^*+i),$$

and minimize it by varying the value β. The interval (0,1] we divide into m equal subintervals and find the value $L(\beta_i)$ at the points $\beta_i=\frac{i}{m}(i=1,2,\ldots,m)$. It is obvious that m gives the accuracy of the finding the minimum of the functional $L(\beta)$. Let $\beta_m^*=\min L(\beta_i)$. Then the forecast of time series we conduct according to the scheme (8), substituting β_m^* for β.

3 Implementation of Forecasting Schemes of Time Series in Artificial Neural Basis

The basis of all forecasting methods is an idea of extrapolation of patterns of the development of the process, which was formed by the time when the forecast came true for future period of time.

Let $v_1,v_2,\ldots,v_t,\ldots.v_n$ is time series. For the synthesis of artificial neural-network forecasting scheme, there must exist a method (methods) of synthesis of neural elements that implement appropriate forecasting models, on whose basis a neural scheme should be constructed. For example, the following artificial neural element with linear activation function implements the autoregression model $\tilde{v}_{n+\tau}=w_1^{(\tau)}v_n+w_2^{(\tau)}v_{n-1}+\ldots\ \ \ldots+w_{k_\tau^*}^{(\tau)}v_{n-k_\tau^*+1}$, with the

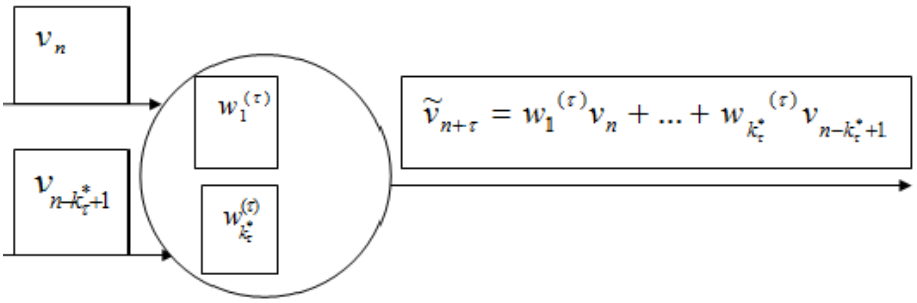


Fig. 1. Neuron of the Optimal Autoregressive Model

optimal step k_τ^* of the prehistory and the step of the forecast τ if $w_1^{(\tau)} = a_1^{*(\tau)}, \ldots w_{k^*}^{(\tau)} = a_{k_\tau^*}^{*(\tau)}$ $a_1^{*(\tau)}, \ldots a_{k_\tau^*}^{*(\tau)}$ are optimal values of parameters of the autoregressive model).

After the development of methods for the synthesis of neural elements that implement the optimal forecasting models in the corresponding classes of models, to predict the values $v_i (i = 1,2,\ldots,n)$ at instants of time $t = n + \tau$, let us design the following neural- network scheme

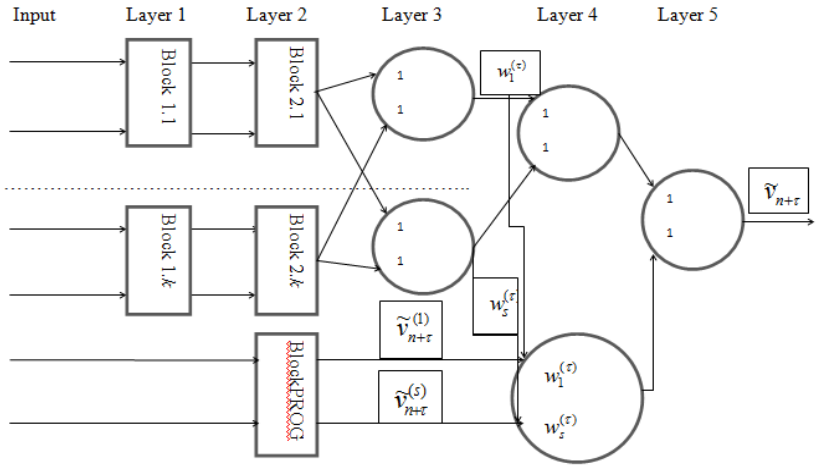


Fig. 2. Neuro-scheme for Time Series Prediction

All the blocks of the 1st layer contain the same number s of neurons, where each neuron implements one of the forecasting models (autoregressive model, polynomial, exponential, linear ones, Brown's linear model, etc.). Neurons that implement the same model in different blocks of this layer have the same serial number.

Each Block 2.m ($m=1,2,....k\,;k=k_\tau^*$) of the 2nd layer contains as much neurons as in Block 1.m. In Block 2.m each neuron has two inputs and a weight vector (1,1), where the value v_{n-k+m} is given to the first input, and the prognostic value $\tilde{v}_{n-k+m,i}^{(\tau)}$ is given to the 2nd input, which is the output signal of the ith neuron of Block 1.m. Activation function of the ith neuron of Block 2.m is set as follows $\exp(-(v_{n-k+m}-\tilde{v}_{n-k+m,i}^{(\tau)})^2)$. The neuron of the serial number i of Block 2.m is related to i th neuron of the 3rd layer in the following way: from the i th neuron of Block 2.m to the m th input of the i th neuron of the 3rd layer there is given the signal $f_{m,i}^{(\tau)}$, where

$$f_{m,i}^{(\tau)}=\begin{cases}1\text{ , if }\ i=\arg\max(\exp(-(v_{n-k+m}-\tilde{v}_{n-k+m,i}^{(\tau)})^2),\\ 0, \qquad\qquad \text{otherwise.}\end{cases}$$

Neurons of the 3rd layer have the linear activation function, and each of the weighting coefficients of each neuron is equal to 1. At the output of the ith neuron of the 3rd layer for the fixed τ we obtain the number $w_i^{(\tau)}$. The 3rd layer, except for neurons with linear activation function, has one more BlokPROG containing exactly as many neurons as a Block of the 1st layer contains. Neurons of this block implement corresponding forecasting model with the depth τ and their serial numbers coincide with the numbers of neurons of Blocks of Layer 1.

The 4th layer contains two linear neurons. The first neuron has s inputs, all its weighting coefficients are equal to 1, and it has activation function $w_1^{(\tau)}+w_2^{(\tau)}+\ldots+w_s^{(\tau)}$.

The second neuron of this layer has weighting coefficients $w_1^{(\tau)},w_2^{(\tau)},\ldots,w_s^{(\tau)}$. If the forecast result of the ith model of BlockPROG is denoted by $\tilde{v}_{n+\tau}^{(i)}$, then at the output of the second neuron of Layer 4 we have $w_1^{(\tau)}\tilde{v}_{n+\tau}^{(1)}+\ldots+w_s^{(\tau)}\tilde{v}_{n+\tau}^{(s)}$.

The 5th layer contains one neuron that has two inputs, a weight vector (1.1), and the activation function $\tilde{v}_{n+\tau}=\dfrac{w_1^{(\tau)}\tilde{v}_{n+\tau}^{(1)}+\ldots+w_s^{(\tau)}\tilde{v}_{n+\tau}^{(s)}}{w_1^{(\tau)}+w_2^{(\tau)}+\ldots+w_s^{(\tau)}}$.

Blocks 2.m ($m=1,2,....k_\tau^*$) determine the most effective basic forecasting models. At the output of the scheme we have a convex linear combination of the best forecasting models.

4 Effectiveness of the Constructed Forecasting Scheme

Following types of errors are often used in the implementation of forecasting time series forecasting

MAE – Mean Absolute Error

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|v_t - \tilde{v}_t\right| \tag{9}$$

where $v_t -$ is the values of the time series at time t;

$\tilde{v}_t -$ predictable value v_t.

The average absolute error of prediction (9) is an absolute measure of the quality of forecast, estimating it independently of the other predictions. It's enough to set a level of absolute error and compare the value of the specified error calculated by the formula (9).

To compare the quality of forecasting, it is often used the average relative error (MRE - Mean Relative Error) is often used

$$MRE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{v_t - \tilde{v}_t}{v_t}\right|, \tag{10}$$

and the average square error (RMSE - Root Mean Square Error) is also used

$$RMRE = \sqrt{\frac{\sum_{t=1}^{n}\left(v_t - \tilde{v}_t\right)^2}{n}}, \tag{11}$$

where v_t are the terms of the time series, $\tilde{v}_t$ are the prognostic values of v_t. RMSE and MRE are relative errors, i.e. they can be used to compare two (or more) different time series prediction the best is the forecast whose value of MRE (10) or RMSE (11) is less.

According to the average relative error criterion, the quality of the forecast of the constructed predicting scheme is estimated by comparing its results with the results of main forecasting models on base of which it is synthesized. To perform this, we use data from the following Table 3 [17].

Table 3. The Original and Forecasted Volumes of Passenger Traffic

Year	Railway	Sea	River	Automobile (coaches)[1]	Aircraft	Under-ground railway
1980	648869	28478.4	24789	7801058	12492.4	430040

1981	653177	30705.6	27531.6	7794859	12720	473437
1982	656485	29362.2	26629.4	7874069	12728.7	515382
1983	668287	29690.2	26810.8	7876161	12711.6	520700
1984	687645	29228.8	24979.6	7998739	12777.8	551851
1985	695129	28660.6	23817.4	8076846	12616	602671
1986	734204	28681	21008.5	8230409	12797.5	598022
1987	717461	27567.3	18750.2	8383820	12670.4	590513
1988	711123	27961.5	20345.5	8552803	13065.3	634616
1989	704078	26524.3	20199.7	8382872	14299.6	648816
1990	668979	26256.7	19090.3	8330512	14833	678197
1991	537407	20786.5	18285.8	7450322	13959.6	595313
1992	555356	13139.5	11158	6464891	5669.3	610668
1993	501495	10497	8064.4	4795664	1947.4	644417
1994	630959	10358.2	6967.9	4039917	1673.3	684480
1995	577432	7817	3594.1	3483173	1914.9	561012
1996	538569	5044.6	2735.9	3304600	1724	536304
1997	500839	4311.3	2443.1	2512147	1484.5	507897
1998	501429	3838.3	2356.5	2403425	1163.9	668456
1999	486810	3084.3	2269.4	2501708	1087	724426
2000	498683	3760.5	2163.3	2557515	1164	753540
2001	467825	5270.8	2034.2	2722002	1289.9	793197
2002	464810	5417.9	2211.9	3069136	1767.5	831040
2003	476742	6929.4	2194.1	3297505	2374.7	872813
2004	452226	9678.4	2140.2	3720326	3228.5	848176
2005	445553	11341.2	2247.6	3836515	3813.1	886598
2006	448422	10901.3	2021.9	3987982	4350.9	917700
2007	447094	7690.8	1851.6	4173034	4928.6	931512
2008	445466	7361.4	1551.8	4369126	6181	958694
2009	425975	6222.5	1511.6	4014035	5131.2	751988
2010	427241	6645.6	985.2	3726289	6106.5	760551
2011	429785	7064.1	962.8	3611830	7504.8	778253
2012	429115	5921	722.7	3450173	8106.3	774058
2013	425217	6642	631.1	3343660	8107.2	774794
2014	**424272.5**	**3490.2**	**453.8915**	**3059461.2**	**9308.8**	**816682.9**
2015	**414375.8**	**5373.2**	**406.4361**	**2645239.9**	**7243.7**	**876984.7**
2016	**425925.3**	**3847.1**	**369.0345**	**2641221.6**	**10609.4**	**972098.3**
2017	**420469.8**	**2975.1**	**233.0464**	**2395820.5**	**10870.7**	**1073108.1**
2018	**426849.1**	**3061.2**	**403.4616**	**2606148.8**	**12330.5**	**1205853.8**

Table 4. Forecast Errors of Passenger Traffic according to MRE criterion

Forecasting methods	Kinds of passenger traffic		
	Railway	River	Automobile
Step of the forecast $\tau = 1$			
Autoregression method	0.0041	0.0148	0.0115
The method of least squares with weights	0.015	0.7975	0.1680
Brown's linear model	0.0358	0.0917	0.1478
Brown's quadratic model	0.0159	0.5516	0.086
Forecasting scheme	0.0039	0.0148	0.0115
Step of the forecast $\tau = 5$			
Autoregression method	0.0045	0.0111	0.0233
The method of least squares with weights	0.0048	0.0683	0.0595
Brown's linear model	0.0585	0.0757	0.1482
Brown's quadratic model	0.0317	0.2295	0.0797
Forecasting scheme	0.0031	0.0108	0.0225

Having analyzed the data in Table 4, we see that the least average relative error occurs in the constructed forecasting scheme. In the two cases (for $\tau = 1$), the error of the scheme coincides with the error of autoregression method. Thus, in general, the scheme developed in this work is the most effective among the methods on which it is based. To obtain the average error (%) of the prediction methods for the given time series in percentage, one should multiply by 100% the corresponding values of quality from Table 4. The quality of the prediction methods of passenger traffic for the forecast period (2014-2018) with the steps of the forecast $\tau = 1$ and $\tau = 5$ is shown in the following charts

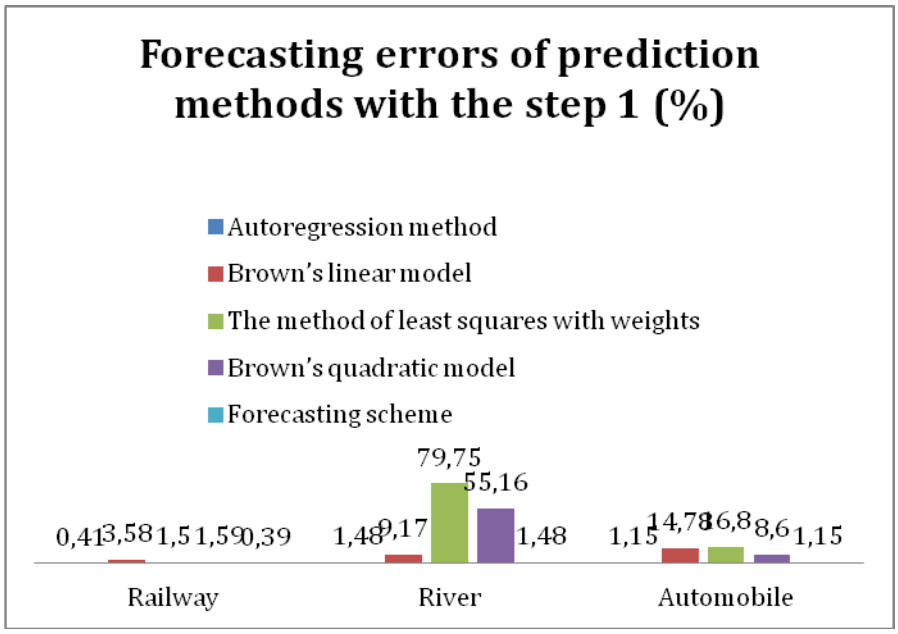


Fig. 3. Forecasting errors of prediction methods with the step 1 (in %)

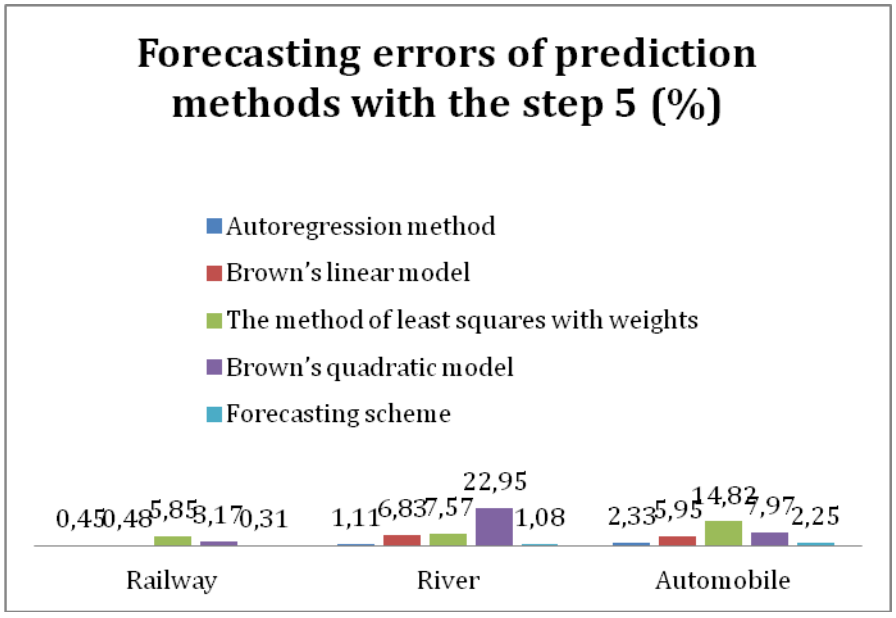


Fig. 4. Forecasting errors of prediction methods with the step 5 (in %)

Note. The constructed forecasting scheme is flexible. This means that a new model can be added to or excluded from basic models (on basis of which the predictive scheme is constructed) at any time. It should be noted that the method of synthesis of the very predictive scheme does not change.

Here are some results of the program implementation of developed forecasting scheme for determining the share of road passenger transport in Ukraine to all other types of transportation during time span since 1980 to 2013. Table 3 contains primary data of passenger traffic volume (period 1980-2013) and projections of passenger traffic (forecast period 2014-2018). On the base of this table it is evident that the average share of road passenger transport in Ukraine was 51.85% over the above mentioned period. Accordingly to the forecast this share will average 45.56% during the prediction period 2014-2018. Thus, the role of road passenger transport in Ukraine over the observable forecast period 2014-2018 is leading. Annual share of road passenger transport in Ukraine during the prediction period is shown on the following diagram:

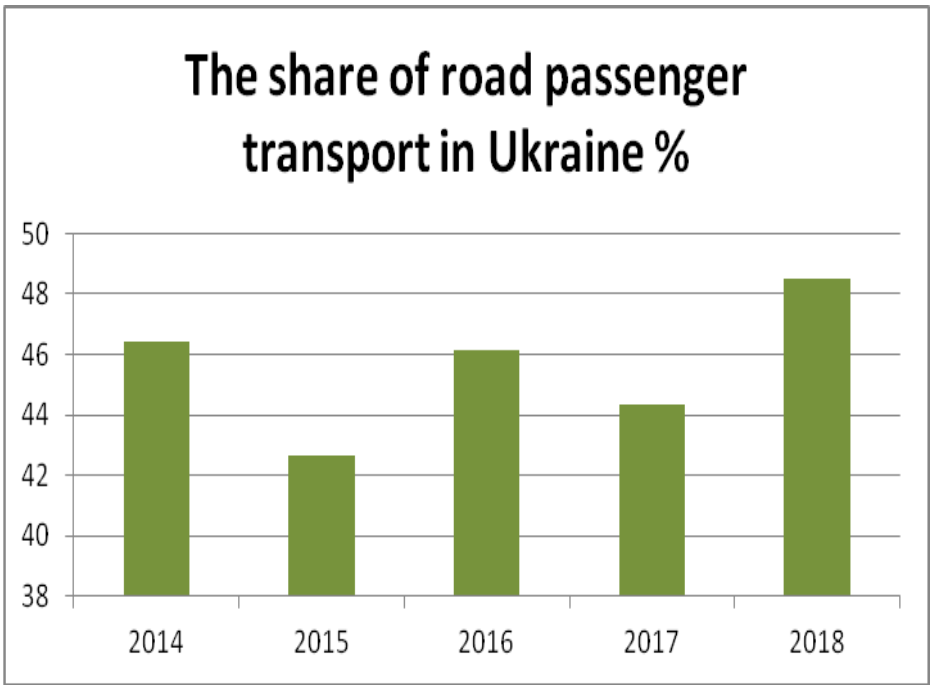
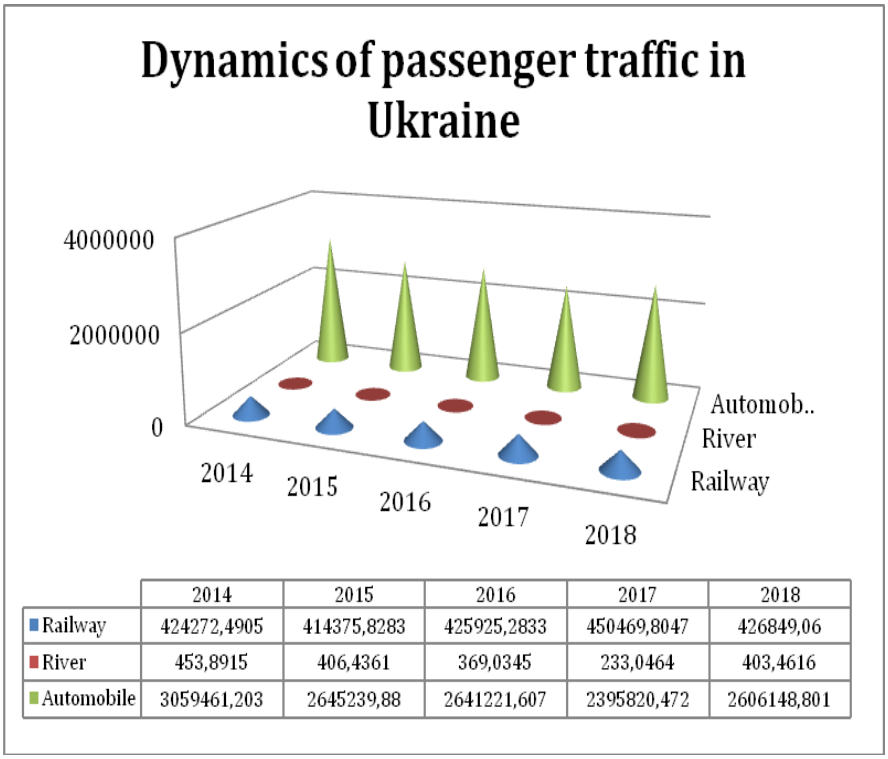


Fig.5. The share of road passenger transport in Ukraine over the period (2014-2018)

To compare the dynamics of changes of the volume of passenger traffic in Ukraine for different types of vehicles (rail, river, road) we construct the following diagram.



	2014	2015	2016	2017	2018
Railway	424272,4905	414375,8283	425925,2833	450469,8047	426849,06
River	453,8915	406,4361	369,0345	233,0464	403,4616
Automobile	3059461,203	2645239,88	2641221,607	2395820,472	2606148,801

Fig.6. Dynamics of passenger traffic in Ukraine (2014-2018)

5 Conclusions

A flexible scheme for forecasting of economic, social, environmental, engineering and technological indicators that can be successfully used in the development of reasonable strategic plans and decisions in the corresponding fields of human activity is worked out.

This forecasting scheme allows us to include new forecasting models of time series or to exclude a model or groups of models from it at any instant of time.

As for the models which remain in the scheme, the competition between them is made over a given period of time, and the final forecasting scheme represents a convex linear combination of models -winners with corresponding weighting coefficients.

References

1. Blank, I.A. Strategy and Tactics of Financial Management. The Item LTD, Kyiv (1996) (in Ukrainian)

2. Heyets, V.M. Iinstability and Economic Growth. Institute of Economic Forecasting of National Academy of Sciences of Ukraine, Kyiv (2002) (in Ukrainian)
3. Zaychenko, Y.P., Moamed, M., Shapovalenko, N.V. Fuzzy Neural Networks and Genetic Algorithms in Problems of Macroeconomic Forecasting. Science news of "Kyiv Polytechnic Institute", 4, 20-30. Kyiv (2002) (in Ukrainian)
4. Ivakhnenko, V. Course of Economic Analysis. Znannya Press, Kyiv (2000). (in Ukrainian)
5. Ivakhnenko, O.H., Lapa, V.G. Prediction of Rrandom Processes. Naukova Dumka, Kyiv (1969) (in Ukrainian)
6. Yarkina, N.M. Econometric Modeling in the Management of Business Risks. Finance of Ukraine, 11, 77-80. Kyiv (2003) (in Ukrainian)
7. Timashova, L., Stepanenko O. Economic-mathematical Evaluation Model of Enterprise in Market Economy. Journal of the Academy of Labour and Social Affairs Federation of Trade Unions of Ukraine, 3 (27), 79-90. Kyiv (2004) (in Ukrainian)
8. Stepanenko, A.P. Modern Computer Tools and Technologies for the Information of the Financial System. New Computer Tools, Computers and Networks, Vol.2, 25-31. Kyiv, Institute of Cybernetics by V.Glushkov of National Academy of Sciences of Ukraine (2001) (in Ukrainian)
9. Tkachenko, R., Pavlyuk, O. Approaches to forecast electricity consumption in power distribution companies // Bulletin "Lviv Polytechnic": Computer Engineering and Information Technology. – № 468. - Pp. 145-151. (2002) (in Ukrainian)
10. Matviichuk, A.V. Modeling of Economic Processes Using Fuzzy Logic Methods. Kyiv National Economic University, Kyiv (2007) (in Ukrainian)
11. Hanke, John E., Arthur, G. Reitsch, and Dean W. Wichern. Business forecasting. Up per Saddle River, NJ: Prentice Hall, (2001)
12. Lewis, Colin David. Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting. Butterworth-Heinemann, (1982)
13. Box, George EP, and Gwilym, M. Jenkins. Time series analysis: forecasting and control, revised ed. Holden-Day, (1976)
14. Tail G. Economic forecasts and decision making / G. Tail. - M .: Statistics. - 448 p. (1971) (in Russian)
15. Kukharev, V.N., Sally V.N., Erpert A.M. Economic-mathematical Methods and Models in the Planning and Management. Vyshcha shcola, Kyiv (1991) (in Russian)
16. Holt, Charles C. "Forecasting seasonals and trends by exponentially weighted moving averages." International Journal of Forecasting 20.1: 5-10. (2004)
17. Winters, Peter R. "Forecasting sales by exponentially weighted moving averages." Management Science 6.3: 324-342. (1960)
18. Transport and Communication in Ukraine - 2013 [Text] / State Statistics Service. Statistical Yearbook, Kyiv (2013) (in Ukrainian).

C-clause calculi and refutation search in first-order classical logic

Alexander Lyaletski

Taras Shevchenko National University of Kyiv, Ukraine
lav@unicyb.kiev.ua

Abstract. The paper describes an approach to the construction of a resolution-type technique basing on a certain generalization of the resolution notion of a clause. This generalization called a conjunctive clause (c-clause) leads to a possibility to introduce two different inference rules and determine two c-clause calculi oriented to refutation search in first-order classical logic both with and without equality. Using the connection of these calculi with Robinson's clash-resolution method, a simple way for the proving of their soundness and completeness is given. Analogs of some of the well-known resolution strategies for the calculi are suggested. Besides, the treatment of Maslov's inverse method in the resolution terms is given. This research can be used in (e-)learning systems for the intelligent testing of knowledge of trainees learning a mathematical subject.

Keywords: first-order classical logic, refutation search, calculus, soundness, completeness, clash-resolution method, paramodulation, strategy

Key Terms: MachineIntelligence

1 Introduction

This paper is devoted to the description of special calculi intended for the establishing of the unsatisfiability of a formula F of a certain form or a set S of such formulas in first-order classical logic maybe with equality. The calculi relate to the class of refutation-search methods based on the ideas firstly presented in Robinsin's paper [1] on the well-known resolution method.

After the appearance of the resolution method, the main efforts of automated theorem-proving community were concentrated on its development in the direction of the construction of its different modifications and strategies oriented to increasing the efficiency of deduction search. All such attempts based on the use of a clause being a well-formed expression of the resolution method leaving aside the possibility of building efficient methods using modifications of the notion of a clause, to which this paper is devoted. Besides, the problem of the interpretation of the Maslov inverse method [2] in resolution terms is solved in it.

Our calculi completely are determined by their (resolution-type) inference rules. The deducibility of a special expression Λ in such a calculus Π is equivalent to the unsatisfiability of F or S. At that, Π is called a *sound* calculus, if the

deducibility of Λ implies the unsatisfiability of F or S; Π is called a *complete* calculus, if the unsatisfiability of F or S implies the deducibility of Λ.

If we put certain restrictions on inferences of Λ in a calculus Π, these restrictions are said to determine a *strategy* for proof search in Π.

All the above-said takes place for the clash-resolution method [3] being called the clause calculus below. It deals with clauses and contains the unique inference rule – the latent class-resolution rule. The empty clause plays the role of Λ.

Let us stop on the way of the construction of an initial set of clauses for a formula F (or a set S of such formulas) being investigated on unsatisfiability. First of all, we can consider that F is a closed formula. Further, let us suppose that F already is presented in Skolem functional form (for satisfiability), all the quantifiers of which are omitted. Then, under the condition that all its variables implicitly are bound by the universal quantifier, the following question is reasonable: Can we refrain from the obligatory presentation F (or S) as a set of clauses and develop a technique similar to the resolution one? Research in this direction is described in what follows. At that, note that the papers [4] and [5] are a starting point for the development of the approach presented here.

We usually give references to the original papers, which laid the foundations for the research in a particular direction, although for the modern description of most of them, one can turn to [6] or [7]. QED indicates the end of any proof.

2 Preliminaries

First-order classical logic with functional symbols and equality is considered.

The notions of terms, atomic formula, and formulas are assumed to be known. A formula being the result of renaming of variables in a formula F is called a *variant* of F. A *literal* is an atomic formula or its negation. For a literal L of the form $\neg A$, its *complementary* $\tilde{L}$ is A. If L is an atomic formula A, then its *complementary* $\tilde{L}$ is $\neg A$.

As it was said above, we restrict ourselves by the consideration of only closed formulas F presented in Skolem functional form for satisfiability by means of the elimination of positive quantifiers. That is, F may be considered as a formula of the form $\forall x_1 \ldots \forall x_m \; G(x_1, , x_m)$, where $x_1, , x_m$ all the variables of F, and $G(x_1, , x_m)$ a quantifier-free formula. I. e. it can be assumed that in the case of reasoning on satisfiability, one has to deal with only quantifier-free formulas, all variables of which implicitly are universally bound.

We can reduce $G(x_1, , x_m)$ to a formula $D_1 \wedge \ldots \wedge D_n$, where D_i is a formula presented in disjunctive normal form (DNF). As a result, we can make investigation of the set $\{D_1, \ldots, D_n\}$ on unsatisfiability instead of making the appropriate investigation of $G(x_1, , x_m)$. This leads to the following notions.

If $L_1, \ldots, L_m$ are literals, then the expression $L_1 \wedge \ldots \wedge L_m$ $(m \geq 1)$ is called a *conjunct*. An expression of the form $C_1 \vee \ldots \vee C_n$, where $C_1, \ldots, C_n$ are conjuncts, is called a *conjunctive clause*, or a *c-clause* $(n \geq 0)$.

A c-clause not containing any conjunct (that is, if $n = 0$) is called an *empty clause* (or *empty c-clause*) and denoted by $\Box$.

In what follows, any conjunct is considered to be the set of its literals and any c-clause – the set of its conjuncts. Thus, in the case when any conjunct of a c-clause contains exactly one literal, this c-clause can be considered as a usual *clause* (see, for example, [1] or [6]).

The introduced definitions allow us to use all the semantic notions of first-order classical logic for c-clauses and sets of c-clauses under the assumption that every variable in any c-clause is universally bound. The empty clause is considered to be an unsatisfiable formula.

Our main purpose is to prove that the inferring of $\square$ in our calculi is equivalent to the unsatisfiability of an initial set of c-clauses.

An *inference* from an initial set S of c-clauses in a calculi under consideration is a sequence $D_1, \ldots, D_n$, where every D_i $(i = 1, \ldots, n)$ is either a variant of an c-clause from S or a variant of a conclusion of a rule applied to some of the c-clauses preceding D_i. Therefore, our calculi *uniquely are identified* by their inference rules. That is why the names of rules will serve as unique names of the calculi under consideration. The *deducibility* of a c-clause C from a set S of c-clauses in a calculus Π is denoted by $S \vdash_{\Pi} C$.

The resolution method first was published in [1] in 1965. It contained the only resolution rule of the arity 2. In [8], J.A.Robinson proposed its modification of this rule under the name of the hyper-resolution. Its further generalization led to the clash-resolution method [3]. The peculiarity of this generalization is that it contains the only latent clash-resolution rule (denoted by RR below) that can be applied to any finite number of clauses. The corresponding clash-resolution method (being the clause calculus with the RR-rule) is sound and complete [3].

Let us give some necessary notations.

A *substitution*, σ, is a finite mapping from variables to terms that has the form $\sigma = \{x_1 \mapsto t_1, \ldots, x_n \mapsto t_n\}$, where variables $x_1, \ldots, x_n$ are pairwise different and for any i $(1 \leq i \leq n)$, the term t_i is distinct from x_i.

A substitution σ is called a *variant substitution* if t_1, ..., t_n from σ are only variables that are pairwise different. In this case, the inverse (one-one) correspondence σ^{-1} exists and presents itself a (variant) substitution.

For an expression Ex and a substitution σ, the result of the application of σ to the expression of Ex is understood in the usual sense; it is denoted by $Ex \cdot \sigma$.

The *composition* of substitutions (as mappings) σ and λ is denoted by $\sigma \cdot \lambda$. It has the property that for any expression Ex, $Ex \cdot (\sigma \cdot \lambda) = (Ex \cdot \sigma) \cdot \lambda$.

For any set Ξ of expressions, $\Xi \cdot \sigma$ denotes the set obtained by the application of σ to each expression in Ξ. If Ξ is a set of (at least two) expressions and $\Xi \cdot \sigma$ a singleton, then σ is called a *unifier* of Ξ. If $\Xi_1, \ldots, \Xi_n$ $(n \geq 1)$ are sets of expressions and for a substitution σ, the set $\Xi_i \cdot \sigma$ is a singleton $(i = 1, \ldots, n)$, then σ is called a *simultaneous unifier* of $\Xi_1, \ldots, \Xi_n$.

It is known (see, for example, [6] or [3]) that in the case the existence of a unifier σ of sets $\Xi_1, \ldots, \Xi_n$, there exist such substitutions λ and σ' that $\Xi_1 \cdot \lambda, \ldots, \Xi_n \cdot \lambda$ are singletons and $\Xi_1 \cdot \sigma = (\Xi_1 \cdot \lambda) \cdot \sigma', \ldots, \Xi_n \cdot \sigma = (\Xi_n \cdot \lambda) \cdot \sigma'$. The substitution λ is unique up to renaming of its variables. It is called the *most general simultaneous unifier (mgsu)* of $\Xi_1, \ldots, \Xi_n$.

Obviously, we can consider that any mgsu σ has the *idempotence property* that means that $\sigma\cdot\sigma=\sigma$. This fact will often be used in what follows implicitly.

Robinson's latent clash-resolution rule (RR). Let clauses $C_0, C_1, \ldots, C_q$ $(q \geq 1)$ with mutually distinct variables be of the forms $C'_0 \vee L_{1,1} \ldots \vee L_{1,r_1} \ldots \vee L_{q,1} \vee \ldots \vee L_{q,r_q}$, $C'_1 \vee E_{1,1} \vee \ldots \vee E_{1,p_1}$, ..., $C'_q \vee E_{q,1} \vee \ldots \vee E_{q,p_q}$ respectively, where $C'_0, C'_1, \ldots, C'_q$ are clauses and $L_1, \ldots, L_q, E_{1,1}, \ldots E_{q,p_q}$ literals. Suppose that there exists the mgsu σ of the sets $\{\tilde{L}_{1,1},\ldots,\tilde{L}_{1,r_1}, E_{1,1},\ldots, E_{1,p_1}\}$, ..., $\{\tilde{L}_{q,1},\ldots, \tilde{L}_{q,r_q}, E_{q,1},\ldots, E_{q,p_q}\}$. Then the clause $C'_0\cdot\sigma \vee C'_1\cdot\sigma \vee \ldots \vee C'_q\cdot\sigma$ is said to be deducible from $C_0, C_1, \ldots, C_q$ by the rule RR.

The RR-rule with two clauses as its premises will be denoted by RR_2.

The paper [3] contains the following result (see, also, [6]).

Robinson's Proposition. An initial set S of clauses is unsatisfiable if and only if the empty clause $\Box$ is inferred in the RR-calculus.

3 C-clause calculi for logic without equality

Below, we introduce two resolution-type rules in order to define two specific c-clause calculi. These calculi have a number of similar properties. That is why their proofs are detailed only for one of them. As to the other calculus, the corresponding proofs for it can be obtained in the same way.

3.1 CR calculus

Let us start with the consideration of the calculus that is based on the analog of Robinson's rule RR.

Clash-resolution (CR). Let c-clauses $D_0, D_1, \ldots, D_q$ $(q \geq 1)$ pairwise without common variables be of the forms $D'_0 \vee K_{1,1} \vee \ldots \vee K_{1,r_1} \vee \ldots \vee K_{q,1} \vee \ldots \vee K_{q,r_q}$, $D'_1 \vee M_{1,1} \vee \ldots \vee M_{1,p_1}$, ..., $D'_q \vee M_{q,1} \vee \ldots \vee M_{q,p_q}$ respectively, where $D'_0, \ldots, D'_q$ are c-clauses and $K_{1,1}, \ldots, K_{q,r_g}$, $M_{1,1}, \ldots M_{q,p_q}$ conjuncts. Suppose that $K_{1,1}, \ldots, K_{q,r_q}$ contain literals $L_{1,1}, \ldots, L_{q,r_q}$ respectively and for every $j = 1, \ldots, q$, $M_{j,1},\ldots M_{j,p_j}$ contain literals $E_{j,1},\ldots E_{j,p_j}$ respectively such that there exists the mgsu σ of the sets $\{\tilde{L}_{1,1},\ldots,\tilde{L}_{1,r_1}, E_{1,1},\ldots, E_{1,p_1}\}$, ..., $\{\tilde{L}_{q,1},\ldots, \tilde{L}_{q,r_q}, E_{q,1},\ldots, E_{q,p_q}\}$. Then the c-clause $D'_0\cdot\sigma \vee D'_1\cdot\sigma \vee \ldots \vee D'_q\cdot\sigma$ is said to *be inferred* from the *nucleus* D_0 and *electrons* $D_1, \ldots, D_q$ by the *CR-rule*. Besides, the q-tuple $\langle D_0, D_1, \ldots, D_q\rangle$ is called a *CR-clash* and $D'_0\cdot\sigma \vee D'_1\cdot\sigma \vee \ldots \vee D'_q\cdot\sigma$ its *CR-resolvent*.

Remark. If $D_0, D_1, \ldots, D_q$ are only clauses, the definitions of CR and RR are coincides, which gives a simple way for proving some results relating to CR.

Proposition 1. *The CR-rule is sound.*

Proof. Since we implicitly consider every variable in any c-clause to be bound by the universal quantifier, obviously it is enough to prove that a CR-resolvent is the logical conclusion of its premises only in the propositional case. For this, it is enough to check the validity of the propositional formula:

$((D'_0 \vee (\tilde{L}_{1,1} \wedge K'_{1,1}) \vee \ldots \vee (\tilde{L}_{1,1} \wedge K'_{1,r_1}) \vee \ldots \vee (\tilde{L}_{q,1} \wedge K'_{q,1}) \vee \ldots \vee (\tilde{L}_{q,1} \wedge K'_{q,r_q})) \wedge (D'_1 \vee (L_{1,1} \wedge M'_{1,1}) \vee \ldots \vee (L_{1,1} \wedge M'_{1,p_1})) \wedge \ldots \wedge (D'_q \vee (L_{q,1} \wedge M'_{q,1}) \vee \ldots \vee (L_{q,1} \wedge M'_{q,p_q}))) \supset (D'_0 \vee D'_1 \ldots \vee D'_q)$, where $\supset$ is the implication symbol, which can be made by applying induction on q. QED.

Let a c-clause D distinguished from $\Box$ be of the form $K_1 \vee \ldots \vee K_n$. Then $\rho(D)$ is the set $\{L_1 \vee \ldots \vee L_n : L_1$ occurs in $K_1, \ldots, L_n$ occurs in $K_n\}$.

For $\Box$, we suppose that $\rho(\Box)$ contains $\Box$ and only it.

If S is a set of c-clauses, then $\rho(S)$ denotes the set $\bigcup_{D \in S} \rho(D)$.

It is obvious that for any non-empty set S of c-clauses, $\rho(S)$ is a finite non-empty set and contains only clauses. Moreover, considering D as a formula, we can produce $\rho(D)$ my means of applying the following propositional tautology: $A \vee (B \wedge C) \equiv (A \vee B) \wedge (A \vee C)$, where $\equiv$ is the logical equivalence symbol. Therefore, *a set S is unsatisfiable if and only if $\rho(S)$ is an unsatisfiable set.*

Remark. According to the previous remark, we conclude that Robinson's clash-resolution technique is used when we are interested in the establishing of the deducibility of $\Box$ from $\rho(S)$ in the CR-calculus. Thus, for any finite set S of c-clauses, it is true that S is unsatisfiable if and only if $\rho(S) \vdash_{CR} \Box$.

Lemma 1. *Let $D_0, D_1, \ldots, D_q$ be c-clauses and $A_0, A_1, \ldots, A_q$ clauses such that $A_0 \in \rho(D_0)$, $A_1 \in \rho(D_1)$, $\ldots$, $A_q \in \rho(D_q)$. If for $A_0, A_1, \ldots, A_q$, there is the CR-clash $\langle A_0, A_1, \ldots, A_q\rangle$ with A_0 as a nucleus and $A_1, \ldots, A_q$ as electrons and A is its CR-resolvent, then there exists the CR-clash $\langle D_0, D_1, \ldots, D_q\rangle$ with D_0 as a nucleus, $D_1, \ldots, D_q$ as electrons, and D as its CR-resolvent such that $A \in \rho(D)$.*

Proof. Let us consider $A_0, A_1, \ldots, A_q$ from the lemma conditions. Let they be of the form: $B_0 \vee L_{1,1} \vee \ldots \vee L_{1,r_1} \vee \ldots \vee L_{q,1} \vee \ldots \vee L_{q,r_q}$, $B_1 \vee E_{1,1} \vee \ldots \vee E_{1,p_1}$, $\ldots$, $B_q \vee L_{q,1} \vee \ldots \vee L_{q,p_q}$ respectively, where B_0, $\ldots$, B_q are clauses and $L_{1,1}$, $\ldots$, L_{q,r_q}, $E_{1,1}$, $\ldots$ E_{q,p_q} literals such that there exists the mgsu σ of the sets $\{\tilde{L}_{1,1}, \ldots, \tilde{L}_{1,r_1}, E_{1,1}, \ldots, E_{1,p_1}\}$, $\ldots$, $\{\tilde{L}_{q,1}, \ldots, \tilde{L}_{q,r_q}, E_{q,1}, \ldots, E_{q,p_q}\}$. Then the clause $B_0 \cdot \sigma \vee B_1 \cdot \sigma \vee \ldots \vee B_n \cdot \sigma$ is a CR-resolvent of the above-given CR-clash.

Since $A_0 \in \rho(D_0)$, D_0 can be presented in the form $D'_0 \vee K_{1,1} \vee \ldots \vee K_{1,r_1} \vee \ldots \vee K_{q,1} \vee \ldots \vee K_{q,r_q}$, where D'_0 is a c-clause and $K_{1,1}$, $\ldots$, K_{q,r_g} conjuncts such that $B_0 \in \rho(D'_0)$ and $K_{1,1}$, $\ldots$, K_{q,r_g} contain literals $L_{1,1}$, $\ldots$, L_{q,r_q} respectively.

Making reasoning in the similar way, we obtain that D_1, $\ldots$, D_q can be presented in the form $D'_1 \vee M_{1,1} \vee \ldots \vee M_{1,p_1}$, $\ldots$, $D'_q \vee M_{q,1} \vee \ldots \vee M_{q,p_q}$ respectively, where D'_1, $\ldots$, D'_q are c-clauses and $M_{1,1}$, $\ldots$ M_{q,p_q} conjuncts such that $B_1 \in \rho(D'_1)$, $\ldots$, $B_q \in \rho(D'_q)$ and $M_{1,1}$, $\ldots$, M_{q,p_q} contain $E_{1,1}$, $\ldots$ E_{q,p_q}.

In accordance with the definition of CR, this means that $D_0, D_1, \ldots, D_q$ form a clash with D_0 as a nucleus and $D_1, \ldots, D_q$ as electrons. For this CR-clash, $D'_0 \cdot \sigma \vee D'_1 \cdot \sigma \vee \ldots \vee D'_q \cdot \sigma$ is its CR-resolvent. Obviously, $B_0 \cdot \sigma \vee B_1 \cdot \sigma \vee \ldots \vee B_n \cdot \sigma \in \rho(D'_0 \cdot \sigma \vee D'_1 \cdot \sigma \vee \ldots \vee D'_q \cdot \sigma)$. QED.

Proposition 2. *Let S be a set of c-clauses and B'_1, $\ldots$, B'_n an inference of $\Box$ from $\rho(S)$ in the RR-calculus. Then there exists an inference B_1, $\ldots$, B_n of $\Box$*

from S in the CR-calculus such that for every j ($j = 1, \ldots, n$) $B'_j \in \rho(B_j)$ and if B'_j is a variant of a CR-resolvent of the CR-clash $\langle B'_{i_r}, \ldots, B'_{i_1}\rangle$ with B'_{i_r} as its nucleus, then B_j is a variant of a CR-resolvent of the CR-clash $\langle B_{i_r}, \ldots, B_{i_1}\rangle$ with B_{i_r} as its nucleus ($i_1 < \ldots < i_r < j$).

Proof. Let $B'_1, \ldots, B'_n$ be an inference of $\Box$ from $\rho(S)$ in the RR-calculus. It is an inference of $\Box$ from $\rho(S)$ in the CR-calculus

For each $i = 1, \ldots, n$, assign a c-clause B_i to a clause B'_i in the following way.

$j = 1$. The definition of an inference implies that B'_1 is a variant of a clause $C \in \rho(S)$. That is there exists a variant substitution λ such that B'_1 is $C \cdot \lambda$. Hence, we can select such a c-clause D in S that $C \in \rho(D)$. Take $D \cdot \lambda$ as B_1. Obviously, $B'_1 \in \rho(B_1)$.

Suppose that $j > 1$ and we have c-clauses $B_1 \ldots, B_{j-1}$ that pairwise have no common variables and satisfy the conditions: $B'_1 \in \rho(B_1), \ldots, B'_{j-1} \in \rho(B_{j-1})$. Two cases are possible.

(1) B'_j is a variant of a clause $C \in \rho(S)$. Proceeding in the same manner as in the case of $j = 1$, we easily achieve the necessary renaming some of the variables of $D \cdot \lambda$ in order the result B_j has no common variables with $B_1, \ldots, B_{j-1}$.

(2) B'_i is a variant of a CR-resolvent C of a CR-clash $\langle B'_{i_r}, \ldots, B'_{i_1}\rangle$ with B'_{i_r} as its nucleus ($i_1 < \ldots < i_r$). Accordantly to Lemma 1, we can construct the CR-clash $\langle B_{i_r}, \ldots, B_{i_1}\rangle$ with B_{i_r} as its nucleus and D as its CR-resolvent, for which $C \in \rho(D)$.

Let λ be a variant substitution such that B'_{i_r} is $C \cdot \lambda$. Obviously, we can select a variant B of $D \cdot \lambda$ not having common variables with $B_1, \ldots, B_{j-1}$ and satisfying the condition $B'_j \in B$. Denote this B by B_j.

Let us consider $B_1, \ldots, B_n$. Since B'_n is $\Box$ and $\rho(\Box)$ contains only $\Box$, B_n is the empty clause $\Box$. Thus, accordingly to the construction of $B_1, \ldots, B_n$, this sequence is an inference of $\Box$ satisfying the conclusion of the proposition. QED.

Now, it is easy to obtain the soundness and completeness of the CR-calculus.

Theorem 1 *(Soundness and completeness of CR-calculus). A set S of c-clauses is unsatisfiable if and only if $S \vdash_{CR} \Box$.*

Proof. The *soundness* of CR is provided by Prop. 1.

Completeness. If S is an unsatisfiable set of c-clauses, then $\rho(S)$ is an unsatisfiable set of clauses. The calculus RR is complete (Robinson's proposition). Hence, $\rho(S) \vdash_{RR} \Box$. Thus, $S \vdash_{CR} \Box$ on the basis of Prop. 2. QED.

Let us consider an example of a deduction in the CR-calculus. Note that all the examples in the paper are given only for propositional case since the resolution-type technique under consideration uses the usual unification.

Example 1. Let U denote the following set of c-clauses: $\{(A \wedge \neg A) \vee (B \wedge C) \vee (E \wedge L),\ \neg B \vee \neg C,\ \neg E \vee \neg L\}$, where A, B, C, E, and L are atomic formulas. The (minimal) inference of $\Box$ from U in CR is as follows:

1. $(A \wedge \neg A) \vee (B \wedge C) \vee (E \wedge L)$ $\quad (\in U)$,
2. $(A \wedge \neg A) \vee (B \wedge C) \vee (E \wedge L)$ $\quad (\in U)$,
3. $(B \wedge C) \vee (E \wedge L)$ $\quad$ (by CR from (1) as a nucleus and (2) as an electron),

4. $(B \wedge C) \vee (E \wedge L)$ (a variant of (3)),
5. $\neg B \vee \neg C$ $(\in U)$,
6. $E \wedge L$ (by CR from (5) as a nucleus and (3) and (4) as electrons),
7. $E \wedge L$ (a variant of (6)),
8. $\neg E \vee \neg L$ $(\in U)$,
9. $\Box$ (by CR from (8) as a nucleus and (6) and (7) as electrons).

Therefore, the set U is unsatisfiable.

3.2 IR calculus

Maslov's inverse method (denoted by MIM here) and Robinson's resolution method (the calculus of clauses in our terminology) appeared approximately at the same time: MIM – in 1964 [2] and RR – in 1965 [1].

After their appearance, the problem of the interpretation of MIM in the resolution terms has arisen. This problem has attracted the attention of a number of researchers in inference search (see, for example, [11] and [12]) also because MIM was defined as a special calculus of so-called favorable assortments and its description was made in the terms that did not correspond to traditional logical terminology and resolution one applied at that time.

In [11], S. Maslov gave himself some MIM explanation in the resolution notions for a restricted case. Later, after an attentive analysis of MIM, the author of this paper "discovered" that MIM interpretation was preferable to do in the terms of a special c-clause[1] calculus [5], the enough description detailed of which is given below. Also it was found that this calculus has an independent significance. It echoes the CR-calculus and, at the same time, it differs from CR.

Inverse resolution (IR). Let c-clauses $D_0, D_1, \ldots, D_q$ $(q \geq 1)$ pairwise without common variables be of the forms $D'_0 \vee K_1 \vee \ldots \vee K_q$, $D'_1 \vee N^1_{1,1} \vee \ldots \vee N^1_{1,p_{1,1}} \vee \ldots \vee N^{r_1}_{1,1} \vee \ldots \vee N^{r_1}_{1,p_{1,r_1}}$, …, $D'_q \vee N^1_{q,1} \vee \ldots \vee N^1_{q,p_{q,1}} \vee \ldots \vee N^{r_q}_{q,1} \vee \ldots \vee N^{r_q}_{q,p_{n,r_n}}$ respectively, where $D'_0, \ldots, D'_q$ are c-clauses and $K_1, \ldots, K_q$, $N^1_{1,1}, \ldots, N^{r_q}_{q,p_{n,r_n}}$ conjuncts. Suppose that for every j $(1 \leq j \leq q)$, K_j contains literals $L_{j,1}$, …, L_{j,r_j} and $N^1_{j,1}$, …, $N^1_{j,p_{j,1}}$, …, $N^{r_j}_{j,1}$, …, $N^{r_j}_{j,p_{j,r_j}}$ contain literals $E^1_{j,1}, \ldots, E^1_{j,p_{j,1}}, \ldots, E^{r_j}_{j,1}, \ldots, E^{r_j}_{j,p_{j,r_j}}$ respectively such that there exists the mgsu σ of the sets $\{\tilde{L}_{1,1}, E^1_{1,1}, \ldots, E^1_{1,p_{1,1}}\}$, …, $\{\tilde{L}_{1,r_1}, E^{r_1}_{1,1}, \ldots, E^{r_1}_{1,p_{1,r_1}}\}$, …, $\{\tilde{L}_{q,1}, E^1_{q,1}, \ldots, E^1_{q,p_{q,1}}\}$, …, $\{\tilde{L}_{q,r_q}, E^{r_q}_{q,1}, \ldots, E^{r_q}_{q,p_{q,r_q}}\}$. Then the c-clause $D'_0 \cdot \sigma \vee D'_1 \cdot \sigma \vee \ldots \vee D'_q \cdot \sigma$ is said to *be inferred* from the *nucleus* D_0 and *electrons* $D_1, \ldots, D_q$ by the IR-*rule*. Besides, the q-tuple $\langle D_0, D_1, \ldots, D_q \rangle$ is called its *IR-clash* and $D'_0 \cdot \sigma \vee D'_1 \cdot \sigma \vee \ldots \vee D'_q \cdot \sigma$ its *IR-resolvent*.

Having the *IR*-rule, we can speak about the IR-calculus.

[1] In 1989, V. Lifschitz independently introducing the notion of a c-clause under the name of a super-clause improved such interpretation [13]. In [14], T. Bollinger extended Loveland's model elimination method [15] to the case of c-clauses using the name of a generalized clause for a c-clause.

The comparative analysis of IR and CR shows that the only difference between them is in the ways of the selection of cutting literals for their applications. The following statement contains a more detailed explanation of this observation.

Lemma 2. *If $\langle D_0, D_1, \ldots, D_q\rangle$ is a CR-clash with D_0 an its nucleus and $D_1, \ldots, D_q$ as its electrons, then for any its CR-resolvent D, it is possible to construct an IR-clash with D_0 as its nucleus and certain variants of $D_1, \ldots, D_q$ as its electrons such that for its some IR-resolvent D' and a substitution τ, $D = D'\cdot\tau$.*

Proof. If $\langle D_0, D_1, \ldots, D_q\rangle$ is the CR-clash from the definition of CR-rule, then the c-clauses $D_0, D_1, \ldots, D_q$ can be presented as $D'_0 \vee K_{1,1} \vee \ldots \vee K_{1,r_1} \vee \ldots \vee K_{q,1} \vee \ldots \vee K_{q,r_q}$, $D'_1 \vee M_{1,1} \vee \ldots \vee M_{1,p_1}$, …, $D'_q \vee M_{q,1} \vee \ldots \vee M_{q,p_q}$ respectively, where $D'_0, \ldots, D'_q$ are c-clauses and $K_{1,1}, \ldots, K_{q,r_g}$, $M_{1,1}, \ldots M_{q,p_q}$ conjuncts and moreover for literals $L_{1,1}, \ldots, L_{q,r_q}$, $E_{j,1}, \ldots E_{j,p_j}$ from $K_{1,1}, \ldots, K_{q,r_g}$, $M_{1,1}, \ldots M_{q,p_q}$ respectively, there exists the mgsu σ of the sets $\Theta_1 = \{\tilde{L}_{1,1},\ldots,\tilde{L}_{1,r_1}, E_{1,1},\ldots, E_{1,p_1}\}$, …, $\Theta_q = \{\tilde{L}_{q,1},\ldots, \tilde{L}_{q,r_q}, E_{q,1},\ldots, E_{q,p_q}\}$ such that $D = D'_0\cdot\sigma \vee D'_1\cdot\sigma \vee \ldots \vee D'_q\cdot\sigma$.

Let us take such variant substitutions $\lambda_{1,1}, \ldots, \lambda_{1,r_1},\ldots, \lambda_{q,1},\ldots, \lambda_{q,r_q}$ that $D_1\cdot\lambda_{1,1},\ldots, D_1\cdot\lambda_{1,r_1}, \ldots, D_q\cdot\lambda_{q,1} \ldots D_q\cdot\lambda_{q,r_q}$ have no common variables with D_0 and each other. Considering $\lambda^{-1}_{1,1}, \ldots, \lambda^{-1}_{q,r_q}$ as mapping graphs, construct the set $\lambda^{-1}_{1,1} \cup \ldots \cup \lambda^{-1}_{q,r_q}$. Obviously, it is a (variant) substitution. Let us denote it by μ and the c-clause $D'_j\cdot\lambda_{j,k} \vee M_{j,1}\cdot\lambda_{j,k} \vee \ldots \vee M_{j,p_j}\cdot\lambda_{j,k}$ by D^k_j.

Let us consider $D^1_1, \ldots, D^{r_1}_1, \ldots, D^1_q, \ldots, D^{r_q}_q$. Accordantly to their definition and the definition of μ, we have that $D^k_j\cdot\mu$ is the same as $D^k_j\cdot\lambda^{-1}_{j,k}$ and, therefore, it is the same as D_j $(j = 1,\ldots,q;\ k = 1,\ldots,r_j)$. Thus, we can select literals $E^1_{1,1}, \ldots, E^1_{1,p_1}, \ldots, E^{r_1}_{1,1}, \ldots, E^{r_1}_{1,p_1}$, …, $E^1_{q,1}, \ldots, E^1_{q,p_q}, \ldots, E^{r_q}_{q,1}, \ldots, E^{r_q}_{q,p_q}$ in $M_{1,1}\cdot\lambda_{1,1}, \ldots, M_{1,p_1}\cdot\lambda_{1,1}, \ldots, M_{1,1}\cdot\lambda_{1,r_1}, \ldots, M_{1,p_1}\cdot\lambda_{1,r_1}, \ldots, M_{q,1}\cdot\lambda_{q,1}, \ldots, M_{q,p_q}\cdot\lambda_{q,1}, \ldots, M_{q,1}\cdot\lambda_{q,r_q}, \ldots, M_{q,p_q}\cdot\lambda_{q,r_q}$ respectively, such that $E^k_{i,j}\cdot\lambda^{-1}_{i,k} = E^k_{i,j}\cdot\mu = E_{i,j}$ $(i = 1,\ldots,q;\ j = 1,\ldots,p_q;\ k = 1,\ldots,r_q)$.

Considering σ and μ as mapping graphs, we conclude that $\zeta = \mu\cdot\sigma \cup \sigma$ is a substitution. Because σ is the mgsu of the sets $\Theta_1,\ldots,\Theta_q$, the definition of ζ and the idempotence of σ imply that ζ is a simultaneous unifier of the sets of literals $\{\tilde{L}_{1,1}\ E^1_{1,1}, \ldots, E^1_{1,p_1}\}$, …, $\{\tilde{L}_{1,r_1}\ E^{r_1}_{1,1}, \ldots, E^{r_1}_{1,p_1}\}$, …, $\{\tilde{L}_{q,1}\ E^1_{q,1}, \ldots, E^1_{q,p_q}\}$, …, $\{\tilde{L}_{q,r_q}\ E^{r_q}_{q,1}, \ldots, E^{r_q}_{j,p_q}\}$. Therefore, there exists the mgsu θ of these sets, for which $\zeta = \theta\cdot\tau$, where τ is a substitution.

As a result, we have that D_0, D^1_1, …, $D^{r_1}_1$, …, D^q_1, …, $D^{r_q}_q$ can form the IR-clash with D_0 as its nucleus and $D^1_1, \ldots, D^{r_1}_1, \ldots, D^q_1, \ldots, D^{r_q}_q$ as its electrons that produces the IR-resolvent $D' = D'_0\cdot\theta \vee D'_1\cdot(\lambda_{1,1}\cdot\theta) \vee \ldots \vee D'_1\cdot(\lambda_{1,r_1}\cdot\theta) \vee \ldots \vee D'_q\cdot(\lambda_{q,1}\cdot\theta) \vee \ldots \vee D'_q\cdot(\lambda_{q,r_q}\cdot\theta)$.

Since $\theta\cdot\tau = \zeta$ and $\zeta = \mu\cdot\sigma \cup \sigma$, it is obvious that $D'\cdot\tau = D$. QED.

This result permits to "simulate" any inference in CR by an inference in IR.

Proposition 3. *Let S be a set of c-clauses and B_1, …, B_n an inference of $\Box$ from S in the CR-calculus. Then there exists an inference B'_1, …, B'_m of $\Box$ from S in the IR-calculus $(m \geq n)$ such that if B_j is a variant of a CR-resolvent of an CR-clash with B_r as its nucleus, then for some j' and r' $(j' \geq j, r' \geq r)$,*

$B'_{j'}$ is a variant of an IR-resolvent of the corresponding IR-clash with $B'_{r'}$ as its nucleus; moreover, $B_j = B'_{j'} \cdot \tau$ for some substitution τ.

Proposition 4. *The IR-rule is sound.*

Proof. As in the case of the CR-rule, it is enough to establish the validity of the following formula, "extracted" from the definition of IR-rule:

$(D'_0 \vee (\tilde{L}_{1,1} \wedge \ldots \wedge \tilde{L}_{1,r_1} \wedge K'_1) \vee \ldots \vee (\tilde{L}_{q,1} \wedge \ldots \wedge \tilde{L}_{q,r_q} \wedge K'_q)) \wedge (D'_1 \vee (L_{1,1} \wedge M^1_{1,1}) \vee \ldots \vee (L_{1,1} \wedge M^1_{1,p_{1,1}}) \vee \ldots \vee (L_{1,r_1} \wedge M^{r_1}_{1,1}) \vee \ldots \vee (L_{1,r_1} \wedge M^{r_1}_{1,p_{1,r_1}})) \wedge \ldots \wedge (D'_q \vee (L_{q,1} \wedge M^1_{q,1}) \vee \ldots \vee (L_{q,1} \wedge M^1_{q,p_{q,1}}) \vee \ldots \vee (L_{q,r_q} \wedge M^{r_q}_{q,1}) \vee \ldots \vee (L_{q,r_q} \wedge M^{r_q}_{q,p_{q,r_q}}))) \supset (D'_0 \vee D'_1 \ldots \vee D'_q)$. QED.

Theorem 2 *(Soundness and completeness of IR-calculus). A set S of c-clauses is unsatisfiable if and only if $S \vdash_{IR} \Box$.*

Proof. The *soundness* is provided by Prop. 4.

Completeness. If S is unsatisfiable set, then $S \vdash_{IR} \Box$ by Theorem 1. By Prop. 3, any inference of $\Box$ from S in CR can be transformed into an inference of $\Box$ from S but already in the IR-calculus, that is $S \vdash_{IR} \Box$. QED.

Example 2. Let us consider the set U from Example 1 and construct the (minimal) inference of $\Box$ from U in IR is as follows:

1. $(A \wedge \neg A) \vee (B \wedge C) \vee (E \wedge L)$ $\quad (\in U)$,
2. $(A \wedge \neg A) \vee (B \wedge C) \vee (E \wedge L)$ $\quad (\in U)$,
3. $(B \wedge C) \vee (E \wedge L)$ (by IR from (1) ac a nucleus and (2) as an electron),
4. $\neg B \vee \neg C$ $\quad (\in U)$,
5. $\neg E \vee \neg L$ $\quad (\in U)$,
6. $\Box$ (by IR from (3) as a nucleus and (4) and (5) as electrons).

We have again proved the unsatisfiability of U.

Draw your attention to the fact that this inference in IR is shorter than the inference in CR from Example 1. This situation is more or less standard for these calculi (see the section containing a comparison of CR and IR).

4 C-clause calculi for logic with equality

The CR- and IR-calculi admit equality handling based on a modification of the paramodulation rule that was proposed in [9] for inference search in first-order theories with equality (denoted by $\simeq$).

We are needed in the following notions that provide us with a possibility to reduce the establishing of the validity of the first-order statement with equality to the search of the refutation of a certain set of c-clauses.

Let S be a set of c-clauses. Then $S^{\simeq}$ denotes the *set of equality axioms* for S in the form of clauses, in which $x, y, z, x_0, \ldots, x_p$ are variables (see, for example, [6]): consists of the following (1) $x \simeq x$, (2) $x \not\simeq y \vee y \simeq x$, (3) $x \not\simeq y \ \vee \ y \not\simeq z \ \vee \ x \simeq z$, (4) $x_i \not\simeq x_o \vee \tilde{R}(x_1, \ldots, x_i, \ldots, x_p) \vee R(x_1, \ldots, x_0, \ldots, x_p)$ for each p-arity predicate symbol R occurring in S and for each $i = 1, 2, \ldots, p$, (5) $x_i \not\simeq x_o$

$\vee\ f(x_1,\ldots,x_i,\ldots,x_p) \simeq f(x_1,\ldots,x_0,\ldots,x_p)$ for each p-arity function symbol f occurring in S and for each $i = 1, 2, \ldots, p$.

A set S of c-clauses is called *equationally unsatisfiable* if and only if the set $S \cup S^{\simeq}$ is unsatisfiable.

Thus, in the case when we have deals with S requiring equality handling, we must establish the equationally unsatisfiability of the set S, which can be achieved by deducing the empty clause $\Box$ from $S \cup S^{\simeq}$. But such approach leads to the extreme large growth of the searching space. For the optimization of such growth, we use a modification of the paramodulation rule [9].

Paramodulation rule PP. Let we have two c-clauses D and $D' \vee (K \wedge s \simeq t)$, where D' is a c-clause and K conjunct (possibly, empty). If there exists mgsu σ of the set of terms $\{s, u\}$, where u is a term occurring in D at a selected position, then the c-clause $D' \cdot \sigma \vee (D \cdot \sigma)[t \cdot \sigma]$ is said to *be inferred* from these c-clauses *by the rule PP*, where $(D \cdot \sigma)[t \cdot \sigma]$ denotes the result of replacing in $D \cdot \sigma$ the term $u \cdot \sigma$ being at the selected position by $t \cdot \sigma$. At that, the ordered pair $\langle\ D,\ D' \vee (K \wedge s \simeq t)\ \rangle$ is called a *PP-clash* (w.r.t. $s \simeq t$) with the *PP-paramodulant* $D' \cdot \sigma \vee (D \cdot \sigma)[t \cdot \sigma]$, *nucleus* D, and *electron* $D' \vee (K \wedge s \simeq t)$.

The *set S^f of functionally reflexive axioms* for a set S of c-clauses consists of all the clauses of the form $f(x_1,\ldots,x_p) \simeq f(x_1,\ldots,x_p)$, where f is a p-arity function symbol occurring in S.

Adding PP to the CR- and IR-calculi, we get the calculi CR+PP and IR+PP intended for inference search in first-order classical logic with equality.

Remark. If in the above-given definition, PP is applied to only clauses, we have the usual paramodulation rule from [9] being denoted by P here.

Because of the completeness of the inference system "negative hyper-resolution + paramodulation" (see, for example, [6]), the following result takes place on the basis that a set S of c-clauses is equationally unsatisfiable if and only if $\rho(S)$ is equationally unsatisfiable.

Robinson-Wos's Proposition. *A set S of c-clauses is equationally unsatisfiable if and only if $\rho(S) \cup \{x = x\} \cup S^f \vdash_{RR+P} \Box$.*

Taking into account the well-known result [10] about the completeness of the system "resolution + paramodulation" without using functionally reflexive axioms, we obtain the further reinforcement of Robinson-Wos's Proposition.

Corollary. *A set S of c-clauses is equationally unsatisfiable if and only if $\rho(S) \cup \{x = x\} \vdash_{RR+P} \Box$. Moreover, RR can denote the only binary rule.*

Now, we have all the necessary for obtaining the results about the completeness of the calculi CR+PP and IR+PP.

First of all, the following analog of Lemma 1 for the PP-rule is obvious.

Lemma 3. *Let D and $D' \vee (K \wedge s \simeq t)$ are c-clauses from the definition of PP. If for $C \in \rho(D)$ and $C' \in \rho(D')$, there exists the PP-clash $\langle C,\ C' \vee s \simeq t\rangle$ w.r.t. $s \simeq t$ with a PP-paramodulant A, then there exists the PP-clash $\langle D,\ D' \vee (K \wedge s \simeq t)\rangle$ w.r.t. $s \simeq t$ with such a PP-paramodulant B that $A \in \rho(B)$.*

Using this lemma and Prop. 2 and 3, it is easy to obtain the following result.

Proposition 5. *Let S be a set of c-clauses and B'_1, ..., B'_n an inference of $\square$ from $\rho(S) \cup \{x = x\} \cup S^f$ in the calculus RR+P. Then there exists an inference B_1, ..., B_n of $\square$ from $S \cup \{x = x\} \cup S^f$ in the calculus CR+PP (IR+PP) such that: (1) if B'_j is a variant of a resolvent of an RR-clash with B'_r as its nucleus, then for some j' and r', $B_{j'}$ is a variant of a resolvent of the corresponding CR-clash (IR-clash) with $B_{r'}$ as its nucleus and, additionally, $B'_r \in \rho(B_{r'} \cdot \tau)$ for some substitution τ; (2) if B'_j is a variant of a paramodulant of a PP-clash with B'_r as its nucleus, then for some j' and r', $B_{j'}$ is a variant of a paramodulant of the PP-clash with $B_{r'}$ as its nucleus and $B'_r \in \rho(B_{r'} \cdot \tau)$ for some substitution τ.*

This proposition, in fact, guarantees the completeness of the *paramodulation extensions* of the CR- and IR-calculi as well as their methods and strategies, some of which are given in the next section. Note that the *soundness* of such extensions is provided by Prop. 1 and 4 and the obvious fact that PP-paramodulant is a logical conclusion of the conjunction of all the c-clauses from $\{N, E\} \cup \{N, E\}^{\simeq}$, where N is a nucleus and E an electron of a PP-rule application.

Theorem 3 *(Soundness and completeness of CR+PP and IR+PP). A set S of c-clauses is equationally unsatisfiable if and only if $S \cup \{x = x\} \vdash_{IR+PP} \square$ ($S \cup \{x = x\} \vdash_{CR+PP} \square$). Moreover, CR (IR) can be the only binary rule.*

Proof. The *soundness* of CR+PP and IR+PP is provided by the remark in the preceding paragraph. *Completeness* takes place for CR+PP and IR+PP due to Corollary and Prop. 5. The completeness of CR+PP with the binary CR-rule is obvious. For proving the completeness of IR+PP with the binary IR-rule, it is enough to note that any binary application of CR can be "decomposed" into r_1 binary applications of IR (see the proof of Lemma 2 for the binary case). QED.

5 Methods and strategies for CR and IR

Prop. 5 gives a simple way for transferring most part of the methods and strategies taking place for the usual clash-resolution (RR) to the ones for the CR- and IR-calculi for classical logic both with and without equality. For the demonstration of how it is possible to do, let us consider the usual liner resolution and positive and negative hyper-resolutions in their wording from [6].

Note that they are given for logic with equality. To obtain them for the case without equality, it is enough to delete all parts concerning the PP-rule in the definitions and wordings of the theorems given below. Also note that their *soundness* is provided by the soundness of the rules CR, IR, and PP. That is why a soundness proof is absent in corresponding theorems.

Linear strategy for CR_2+PP and IR_2+PP. It permits to apply CR_2 (IR_2) or PP to the pair of c-clauses when beginning with the second rule application in an inference, any its c-clause is either a CR_2-resolvent (IR_2-resolvent) or PP-paramodulant of the previous application of the rule CR_2 (IR_2) or PP, and the other c-clause is a variant of either a c-clause from an initial set S of

c-clauses or a c-clause that was deduced earlier.

Theorem 4 *(Soundness and completeness of linear strategy for CR_2+PP and IR_2+PP). A set S of c-clauses is equationally unsatisfiable if and only if there exists an inference of $\square$ from $S \cup \{x = x\} \cup S^f$ satisfying to the linear strategy for CR_2+PP (IR_2+PP).*

Proof. Completeness takes place due to the completeness of the usual linear resolution with paramodulation [6], Robinson-Wos's Proposition, and Prop. 5. QED.

Positive and negative hyper-resolution for CR_2+PP (IR_2+PP). An atomic formula is called a *positive* literal. A literal of the form $\neg A$, where A is an atomic formula, is a *negative* one.

A c-clause is called a *positive* (*negative*) if each its conjunct contains at least one positive (negative) literal. Note that there are c-clauses being positive and negative at the same time, for example, $\neg A \wedge A$.

A CR- or IR-clash $\langle D_0, D_1, \ldots, D_q \rangle$ with D_0 as a nucleus and $D_1, \ldots, D_q$ as electrons is called *positive* (*negative*), if $D_1, \ldots, D_q$ are positive (negative) c-clauses and the cut literals $L_{j,k}$ in the definitions of CR or IR respectively are negative (positive).

For logic without equality, *positive (negative) hyper-resolution strategy* for CR and IR permits constructing inferences containing only the positive (negative) hyper-resolution clashes with positive (negative) CR- or IR-resolvents.

In the case of logic with equality, we additionally permit to apply the PP-rule only to positive nucleus and electron; moreover, a literal containing the selected occurrence of the term u (see the definition of PP-rule) must be positive.

Theorem 5 *(Soundness and completeness of positive and negative hyper-resolutions with PP-rule). A set S of c-clauses is equationally unsatisfiable if and only if there exists an inference of $\square$ from $S \cup \{x = x\} \cup S^f$ satisfying to the positive and negative hyper-resolution with CR- (IR-) and PP-rules.*

Proof. Completeness. Since there exists an inference of $\square$ from $\rho(S) \cup \{x = x\} \cup S^f$ satisfying to the usual positive (negative) hyper-resolution and paramodulation (see [6]), this inference can be transformed into an inference of $\square$ from $S \cup \{x = x\} \cup S^f$ satisfying to the positive and negative hyper-resolution with CR (IR) and PP on the basis of Robinson-Wos's Proposition and Prop. 5. QED.

Remark. In Theorems 9 and 10, the adding of functionally reflexive axioms to the set S is the necessary condition for completeness. Examples demonstrating this for clauses (when CR, IR, and RR are coincided) can be found in [6].

6 IR calculus and Maslov's inverse method

Below, we give the description of MIM in the form of a special strategy for IR.

Maslov's inverse method deals with so-called favorable assortments. In this connection, we consider MIM as a calculus of favorable assortments that has two inference rule: A and B. The A rule determines an initial set of favorable

assortments, while the B rule produces new favorable assortments from the already deduced ones. That is why we treat assortments as clauses and favorable assortments as favorable clauses being produced by the α and β rules (see below).

If C is a conjunct $L_1 \wedge \ldots \wedge L_r$, where $L_1, \ldots, L_r$ are literals, then $\tilde{C}$ denotes the clause $\tilde{L}_1 \vee \ldots \vee \tilde{L}_r$.

Rule α. Let S be a set of c-clauses and $S^d = \{\tilde{C} : C$ is a conjunct from a c-clause belonging to $S\}$. If $S^\alpha = \{C : C = C' \cdot \sigma \vee C'' \cdot \sigma$, where C', $C'' \in S^d$ and C' and C'' contain literals L and L' respectively such that there exists the mgsu σ of $\{\tilde{L}, L'\}\}$, then any clause from S^α is called a $favorable$ one deduced from S by the α-$rule$.

Obviously, S^α is a finite set if S is the same. Besides, each its (favorable) clause contains both a literal and its complementary. That is why S is a unsatisfiable set of c-clauses if and only if the set $S \cup S^\alpha$ is unsatisfiable.

Rule β. Let S be a set of c-clauses, $D \in S$, D consists of q conjuncts, and C_1, …, C_q be favorable clauses. If the IR-rule can be applied to D as a nucleus and C_1, …, C_q as electrons, than the IR-resolvent of this application is called a $favorable$ clause that is deducible from D, C_1, …, C_q by the β-$rule$.

Note that the requirement that the number of conjuncts in D ids equal to q leads to the fact that any IR-resolvent of β-rule is a clause.

In these terms, MIM presents itself the following strategy for IR-calculus called a MIM-*strategy*: First of all, we produce all the possible favorable clauses applying the α-rule; then, we apply only the β-rule attempting to deduce $\square$.

The soundness of the MIM-strategy provides the soundness of IR-rule and the above-given remark about $S \cup S^\alpha$. As to completeness, the proof of it is omitted here; we simply give the rewording of the main result for MIM from [2].

Theorem 6 *(Soundness and completeness of MIM-strategy). A set S of c-clauses that pairwise have no common variables is unsatisfiable if and only if there exists an inference of $\square$ from S satisfying to the MIM-strategy.*

This result seams unexpected because of the requirement that D from the definition of the β-rule must consist of exact q conjuncts. This apparent contradiction is explained by the fact that when using the MIM-strategy, we construct S^α containing clauses, the usage of which in an application of the β-rule can be considered as a "latent" way for reducing the number of electrons.

The below-given example demonstrates some of the features of inferences satisfying to the MIM-strategy.

Example 3. It is easy to see that for U from Example 1, $U^d = \{\neg A \vee A, \neg B \vee \neg C, \neg E \vee \neg L, B, C, E, L\}$. As a result, $U^\alpha = \{\neg A \vee A \vee \neg A \vee A, \neg B \vee \neg C \vee B, \neg B \vee \neg C \vee C, \neg E \vee \neg L \vee E, \neg E \vee \neg L \vee L\}$. We have the following MIM-inference:

1. $(A \wedge \neg A) \vee (B \wedge C) \vee (E \wedge L)$ $(\in U)$,
2. $\neg B \vee \neg C$ $(\in U)$,
3. $\neg E \vee \neg L$ $(\in U)$,
4. $\neg A \vee A \vee \neg A \vee A$ (by α-rule),

5. $\neg B \vee \neg C \vee B$ (by α-rule),
6. $\neg B \vee \neg C \vee C$ (by α-rule),
7. $\neg E \vee \neg L \vee E$ (by α-rule),
8. $\neg E \vee \neg L \vee L$ (by α-rule),
9. $B \vee E$ (by β-rule from (1) as a nucleus and (4), (5), and (7) as electrons),
10 . $C \vee E$ (by β-rule from (1) as a nucleus and (4), (6), and (7) as electrons),
11. E (by β-rule from (2) as a nucleus and (9) and (10) as electrons),
12. $B \vee L$ (by β-rule from (1) as a nucleus and (4), (5), and (8) as electrons),
13 . $C \vee L$ (by β-rule from (1) as a nucleus and (4), (6), and (8) as electrons),
14. L (by β-rule from (2) as a nucleus and (12) and (13) as electrons),
15. $\square$ (by β-rule from (3) as a nucleus and (11) and (14) as electrons).

We have proved the unsatisfiability of U at the 3rd time.

7 Comparison of CR- and IR-calculi

One can see that the obtained results on the CR- and IR-calculi "echo" each other. In this connection, it is interesting to know is there any advantages of one of them over the other? Moreover that Prop. 3 states that any inference of $\square$ in CR can be simulated by an inference of $\square$ in IR with the same number of rule applications. This section contains an answer on this question when comparison is made w.r.t. inferences being minimal on the number of rule applications.

By $\psi(\Pi, \Delta, S)$, denote the number all the c-clauses in an inference Δ of a c-clause C from a set S in a calculus Π that are deduced by different rule applications. The inference Δ is *minimal on the number of rule applications* if for any other inference Δ' of a variant of C from S in Π, the inequality $\psi(\Pi, \Delta, S) \leq \psi(\Pi, \Delta', S)$ holds.

Let Δ denote an inference of $\square$ from S in CR. Using Prop. 3, it is easy to construct an inference Γ of $\square$ from S in IR such that $\psi(IR, \Gamma, S) \leq \psi(CR, \Delta, S)$. Thus, in the case when Δ_{min} and Γ_{min} denotes the minimal inferences on the introduced characteristic, we have that $\psi(CR, \Delta_{min}, S) - \psi(IR, \Gamma_{min}, S) \geq 0$.

Let us make an attempt to find an upper bound for this difference restricting us by the case when an initial set S contains only c-clauses without variables.

Let us consider an application of IR-rule to a nucleus c-clause D_0 and electron clauses $D_1, \ldots, D_n$ $(n \geq n)$ with an IR-resolvent D. Its attentive analysis demonstrates that this $(n+1)$-arity application can be slitted into n binary applications of CR-rule in the following way: first we make a binary application of CR to D_0 and D_1, then to an obtained CR-resolvent and D_2, and so on. That is we can split any $n+1$-arity IR-application into n binary CR-applications in such a way that for the result C of such CR-rule applications, C will contain all or some of conjuncts belonging to D.

This observation leads to the following upper bound for the difference given above: $\psi(CR, \Delta_{min}, S) - \psi(IR, \Gamma_{min}, S) \leq \sum(m_i - 2)$, where m_i is the arity of the ith CR-rule application in Δ_{min} and the sum is taken over all of m_i.

To demonstrate that this upper bound is achieved, let us take the sets $S_n = \{(L_1 \wedge E_1) \vee \ldots \vee (L_n \wedge E_n), (A_{1,1} \wedge B_{1,1}) \vee \ldots \vee (A_{1,m_1} \wedge B_{1,m_1}) \vee \tilde{L}_1 \vee$

$\tilde{E}_1$, $\tilde{A}_{1,1} \vee \tilde{B}_{1,1} \vee \tilde{L}_1 \vee \tilde{E}_1$, ..., $\tilde{A}_{1,m_1} \vee \tilde{B}_{1,m_1} \vee \tilde{L}_1 \vee \tilde{E}_1$, ..., $(A_{n,1} \wedge B_{n,1}) \vee \ldots \vee (A_{n,m_n} \wedge B_{n,m_n}) \vee \tilde{L}_n \vee \tilde{E}_n$, $\tilde{A}_{n,1} \vee \tilde{B}_{n,1} \vee \tilde{L}_n \vee \tilde{E}_n$, ..., $\tilde{A}_{n,m_n} \vee \tilde{B}_{n,m_n} \vee \tilde{L}_n \vee \tilde{E}_n\}$, where L_1, ..., L_n, E_1, ..., E_n, $A_{1,1}$, ..., A_{n,m_n}, $B_{1,1}$, ..., B_{n,m_n} are literals.

Below we give an inference $\overline{\Delta}$ of $\Box$ from S_n in the IR-calculus. (Thus, S_n is an unsatisfiable set.)

$\lceil (A_{1,1} \wedge B_{1,1}) \vee \ldots \vee (A_{1,m_1} \wedge B_{1,m_1}) \vee \tilde{L}_1 \vee \tilde{E}_1 \quad (\in \overline{S})$,
$\mid \tilde{A}_{1,1} \vee \tilde{B}_{1,1} \vee \tilde{L}_1 \vee \tilde{E}_1 \quad (\in \overline{S})$,
$\mid \quad \ldots$
$\lfloor \tilde{A}_{1,m_1} \vee \tilde{B}_{1,m_1} \vee \tilde{L}_1 \vee \tilde{E}_1 \quad (\in \overline{S})$,
$\quad \ldots$
$\lceil (A_{n,1} \wedge B_{n,1}) \vee \ldots \vee (A_{n,m_n} \wedge B_{n,m_n}) \vee \tilde{L}_n \vee \tilde{E}_n \quad (\in \overline{S})$,
$\mid \tilde{A}_{n,1} \vee \tilde{B}_{n,1} \vee \tilde{L}_n \vee \tilde{E}_n \quad (\in \overline{S})$,
$\mid \quad \ldots$
$\lfloor \tilde{A}_{n,m_n} \vee \tilde{B}_{n,m_n} \vee \tilde{L}_n \vee \tilde{E}_n \quad (\in \overline{S})$,
$\lceil (L_1 \wedge E_1) \vee \ldots \vee (L_n \wedge E_n) \quad (\in \overline{S})$,
$\mid \tilde{L}_1 \vee \tilde{E}_1$ (by IR from the 1st-block c-clauses with the 1st c-clause as a nucleus),
$\mid \quad \ldots$
$\lfloor \tilde{L}_n \vee \tilde{E}_n$ (by IR from the nst-block c-clauses with the 1st c-clause as a nucleus),
$\Box$ (by IR from the $(n+1)$st block c-clauses with the 1st c-clause as a nucleus).

Using the ideas from [16], we can prove that $\overline{\Delta}$ is a minimal inference in IR containing $n+1$ rule applications with the arities $m_1+1, \ldots, m_n+1$, and $n+1$.

Now, let us convert $\overline{\Delta}$ into an inference $\overline{\Gamma}$ of $\Box$ from S_n, but already in the CR-calculus in the following way:

For each i $(i = 1, \ldots, n)$, let us replace the c-clause $\tilde{L}_i \vee \tilde{E}_i$ by the sequence of c-clauses $(A_{i,2} \wedge B_{i,2}) \vee \ldots \vee (A_{i,m_i} \wedge B_{i,m_i}) \vee \tilde{L}_i \vee \tilde{E}_i$, ..., $(A_{i,m_i} \wedge B_{i,m_i}) \vee \tilde{L}_i \vee \tilde{E}_i$, $\tilde{L}_i \vee \tilde{E}_i$ that along with the all c-clauses form the i th block is an inference of $\tilde{L}_i \vee \tilde{E}_i$ in CR. Replace the empty clause $\Box$ by the sequence $(L_2 \wedge E_2) \vee \ldots \vee (L_n \wedge E_n)$, ..., ... $(L_n \wedge E_n)$, $\Box$, being an inference of $\Box$ in CR since $(L_2 \wedge E_2) \vee \ldots \vee (L_n \wedge E_n)$ is deduced from $(L_1 \wedge E_1) \vee (L_2 \wedge E_2) \vee \ldots \vee (L_n \wedge E_n)$ and $(L_1 \wedge E_1)$ by the CR-rule, ..., $(L_n \wedge E_n)$ is deduced from $(L_{n-1} \wedge E_{n-1}) \vee \ldots \vee (L_n \wedge E_n)$ and $(L_{n-1} \wedge E_{n-1})$ by the CR-rule, $\Box$ is deduced from $(L_n \wedge E_n)$ and $(L_n \wedge E_n)$ by CR.

We have that $\overline{\Gamma}$ is an inference of $\Box$ from S_n in CR, for which $\psi(CR, \overline{\Gamma}, S_n) = (\sum_{i=1}^{n} m_i) + (n+1)$. Again using the ideas from [16], we can conclude that $\overline{\Gamma}$ is a minimal inference in CR.

Finally, we get $\psi(CR, \overline{\Gamma}, S_n) - \psi(IR, \overline{\Delta}, S_n) = (n-1) + \sum_{i=1}^{n}(m_i - 1)$, that is the upper bound is reachable.

8 Conclusion

The paper does not touch any practical aspects and is purely theoretical. Nevertheless, the author considers that it may be useful for researchers involved in the implementation of intelligent systems, in particular, e-learning systems requiring tools for proof search in classical logic at least for the following reasons.

The research demonstrates that the transition to c-clauses being the generalization of the widely-used resolution notion as a clause gave the possibility to construct the calculi possessing different properties in general and not worsening such an important characteristic as the minimum number of rule applications in comparison with the usual resolution methods. Although now it is difficult to say that the "behavior" of provers based on these calculi will be better than the "behavior" of the well-know resolution provers such as Vampire or Prover 9, we may expect that more detailed analysis of the proposed approach will lead to the further improvement of the traditional resolution technique. From this point of view, MIM seems to be a more attractive method, possessing a number of positive features not mentioned in the paper and requiring a separate study.

References

1. J. A. Robinson. A machine-oriented logic based on the resolution principle. In *J. Assoc. Comput. Mach.* 12, 23-41, 1965, 28: 2–20.
2. S. Yu. Maslov. The inverse method for establishing the deducibility in the classical predicate calculus. In *DAN SSSR*, 159(1): 17–20, 1964. In Russian.
3. J. A. Robinson. An Overview of mechanical theorem proving. In *Lecture Notes in Operations Research and Mathematical Systems*, 28: 2–20, 1970.
4. A. V. Lyaletski and A. I. Malashonok. A calculus of c-clauses based on the clash-resolution rule. In *Mathematical Issues of Intellectual Machines Theory*, GIC AS UkrSSR: Kiev, 3–33, 1975. In Russian.
5. A. V. Lyaletski. On a calculus of c-clauses. In *Mathematical Issues of Intellectual Machines Theory*, GIC AS UkrSSR: Kiev, 34–48, 1975. In Russian.
6. Ch. Lee and R. Ch. Chang, Richard (1987). *Symbolic Logic and Mechanical Theorem Proving*. Academic Press: New York, 331 pp., 1997.
7. J. A. Robinson and A. Voronkov, editors. *Handbook of Automated Reasoning* (volume 1). Elsevier and MIT Press, 981 pp., 2001.
8. J. A. Robinson. Automatic deduction with hyper-resolution. In *International Journal of Computer Mathematics*, 227–234, 1965.
9. G. Robinson and L. Wos. Paramodulation and theorem-proving in first-order theories with equality. In *Machine Intelligence*, 4: 135–150, 1969
10. D. Brand. Proving theorems with the modification method. In *SIAM Journal on Computing*, 4: 412–430, 1975.
11. S. Yu. Maslov. Proof-search strategies for methods of resolution type. In *Machine Intelligence*, 6: 77–90, 1971.
12. D. Kuechner. On the relation between resolution and Maslov's inverse method. In *Machine Intelligence*, 6: 73–76, 1971.
13. Lifschitz V. What is the inverse method? In *Journal of Automated Reasoning*, 5: 1–23, 1989.
14. Bollinger T. A model elimination calculus for generalized clauses. In *Proceedings of IJCAI'91*, v. 1: 126–131 , 1991.
15. Loveland D.W. A simplified format for the model elimination theorem-proving procedure. In *Journal of the ACM (JACM)*, v. 16, n. 3: 349–363, 1969.
16. A. V. Lyaletski. On minimal inferences in the calculi of c-clauses. In *Issues of the Theory of Robots and Artificial Intelligence*, GIC AS UkrSSR: Kiev, 88–101, 1977. In Russian.

Principles of intellectual control and classification optimization in conditions of technological processes of beneficiation complexes

Andrey Kupin[1] and Anton Senko[1]

[1] Department of Computer Systems and Networks, Faculty of Information Technologies,
Kryviy Rih National University, Partzyizdu str., 11,
50027 Kryviy Rih, Ukraine
kupin@mail.ru, antonysenko@gmail.com

Abstract. These theses contains realization of a typical technological beneficiation complex for automation of control processes (in the context of beneficiation of iron ore - magnetite quartzites). The hierarchy scheme of intelligence control system for such complex combining principles of neurocontrol, classification and optimal control has been shown. Results of computer modeling of classification optimization process in the context of actual indicators of magnetite quartzites concentration have been shown.

Keywords. Intellectual control, classification optimization, beneficiation technology, iron ore, magnetite quartzites

Key Terms. Intelligence, Control System, Model, Classification

1 Introduction

Nowadays the problem of intellectual control of technological processes is considered rather actual. Thus necessity of constant improvement of manufacture, increase of competitiveness, minimization of technological environmental impact demands application of complex automation systems is based on modern information technologies (IT) and intelligent control systems (ICS) [1].

Let's consider the complex of technological processes of iron ore beneficiation (magnetite quartzites). As the object of control such complex is characterized by sufficient complexity (multichanneling, nonlinearity, non-stationary, illegibility and incompleteness of information along with great value of transport delay of output parameters, presence of noise and disturbance, presence of recycles on the majority of stages, etc.) [2]. Taking into account these properties, statement of a problem and

potential approaches to their decision such complex can be considered as typical [3-4].

Works of [2-10] are of great importance for the development of intellectual control theory of beneficiation technology objects. At the same time, despite of considerable quantity of research and development, existing systems of automation do not always meet modern requirements and do not provide the effective decision of difficult tasks in actual conditions in beneficiation process line.

2 Review of existing decisions and task setting

Taking into account multidimensionality, illegibility and incompleteness of technological information on all levels of control it is necessary to use ICS to support operators' (controllers, technologists and other) decision making and increase their quality [1]. The further task setting of intellectual control of a process line (a section) can be also conditionally represented by means of classical cybernetics chart "black box" (Fig. 1). Accordingly, for controlling the beneficiation process set of vectors *X, U, Y, V* on the basis of can be formed as follows.

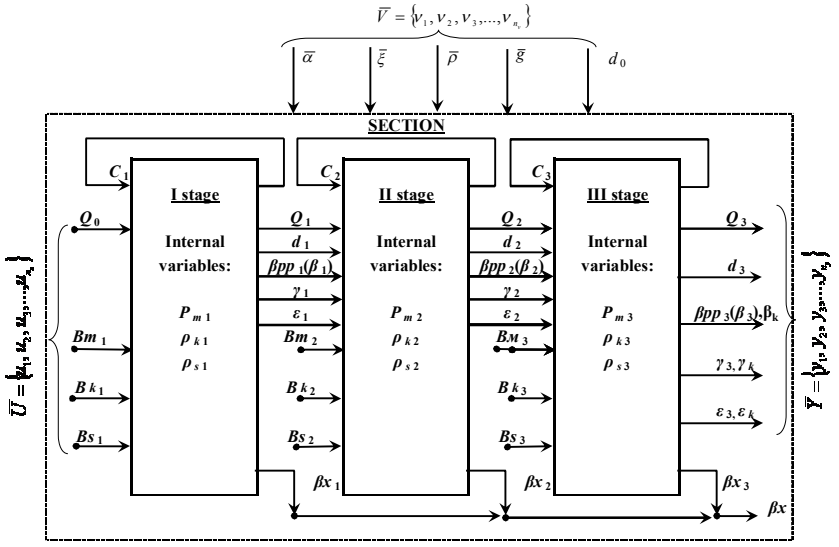


Fig. 1. Process line (section) of concentrating as the object of intelligence control

In Fig. 1 such notations are taken: $i = 1...N_r$ is a number of industrial variety of ore; N_r is quantity of industrial varieties; $\overline{\alpha} = \{\alpha_i\}$ is estimated raw ore grade; $\overline{\xi} = \{\xi_i\}$ is specific gravity of every variety of ore; $\overline{\rho} = \{\rho_i\}$ is an index or a group of indices that characterize physical and chemical properties of ore (for example, density of corresponding varieties of ore, strength, grindability, etc.); $\overline{g} = \{g_i\}$ is index that

characterizes mineralogical and/or morphological properties of ore (for example, averaged size of magnetite dissemination in ore after varieties); d_0 is averaged ore coarseness before beneficiation; Q_0 is an ore consumption on the first stage of beneficiation; $j = 1...N_s$ is number of beneficiation stage; N_s – is quantity of stage; $\overline{Q} = \{Q_j\}$, is processing output of each stage; $\overline{C} = \{C_j\}$ is circulation load; $\overline{d} = \{d_j\}$ is averaged product coarseness; $\overline{P}_m = \{P_{m_j}\}$ is a solid content in pulp; $\overline{B}_m = \{B_{m_j}\}, \overline{B}_k = \{B_{k_j}\}, \overline{B}_s = \{B_{s_j}\}$ are consumption of water to the mill, classifier and magnetic separation respectively; $\overline{\rho}_k = \{\rho_{k_j}\}$ is a pulp density in the process of classification; $\overline{\rho}_s = \{\rho_{p_j}\}$ is a pulp density before magnetic separation; $\overline{\beta}_{pp} = \{\beta_{pp_j}\} = \{\beta_j\}$ is an estimated grade in the industrial product; $\overline{\beta}_x = \{\beta_{x_j}\}$ is loss of a commercial component in tails; β_k is a quality of concentrate; $\overline{\gamma} = \{\gamma_j\}$ is an output of useful component in an industrial product; γ_k is an output of useful component in concentrate; $\overline{\varepsilon} = \{\varepsilon_j\}$ is an extraction of useful component in an industrial product; ε_k is an extraction of useful component in a concentrate.

Thus distribution of state vector on input and output indexes is conditional enough because most parameters on output, for example, of the first stage will be input for the second, etc.

For further application of multidimensional model such as Fig. 1 (for example, for decision of identification tasks or synthesis of automated control systems of beneficiation TP) with using artificial intelligence technology a number of typical neural network structures that will be offer by the author here.

3 The hierarchy scheme of intelligence control system for such complex combining principles of neurocontrol, classification and optimal control

The results of tests of such intelligent systems have proved the possibility of their application in the beneficiation TP. At the same time, to ensure their operation it is necessary to determine the values of settings and / or trends in their paths. Further studies have shown that the determination of the required setting values it is necessary to carry out by combination of the following [7]:

1. Classification control, that is founded on the basis of permanent accumulation of technological parameters history database (DB), their grouping on certain signs (clustering) and determination of value of setting for the measure of similarity to the current values of vectors: input, output and internal parameters[8, 9].

2. Optimal control, which requires the design of general purpose functionality for the system and the application of global optimization methods [4, 10].

Main advantages of the classification approach are their potentially high fast-acting due to the use of well-known methods of clustering and patterns recognition (for example, neural networks classification). The disadvantage is low accuracy (the

chosen decision is not necessarily optimal, and even quasioptimal). Also, application of the approach does not always guarantee the result. In particular, this may be due to such cases:

- at the beginning of the system operation, when the database of technological situations parameters is quite small;
- in the case when necessary (similar) combination of parameters (cluster) has not been met yet in the process of exploitation of ICS;
- in changing of flowsheet, regime map, presence of considerable disturbance of properties of primary raw material (ore, its amount and correlation of mineral varieties, etc.).

On the one side, optimization approaches in the case of multidimensional goal function are also characterized by disadvantages that are caused by:

- the difficulty of obtaining a sufficiently adequate mathematical model of TP [4], which is typical for most inertial processes (in particular, the beneficiation);
- the bad conditionality of optimization task (presence of great amount of local extremums) that appears in the case of application of well-known identification methods of the multidimensional systems (regressive models, Wiener–Hopf equation, synergetic and self-organizations, artificial neural networks and others in particular) and greatly limits the application of well-known methods of multidimensional optimization;
- slow convergence rate of computing process during optimization in large number of cases.

On the other hand, in the case of the possibility of designing the mathematical model and a good choice of hill climbing algorithm (method) it is possible to solve control task, which allows to define a really optimal (or quasioptimal) settings, with certain limitations. Taking into account well-known advantages and disadvantages of the above-mentioned approaches for the implementation of multichannel ICS of TP of iron-ore beneficiation the approach based on combination of classification and optimization algorithms has been offered. Structure of multichannel hierarchical ICS of TP of beneficiation complex based on the system of coupling of neurocontrol, classification and optimization methods is shown in Fig. 2.

In Fig. 2 such notations are taken: OC_{ij} is a control object (channel), j its number (j=1,…,k_i; k_i is an amount of control channels), i is a number of the stage for local TP (for example, fragmentation, classification, magnetic separation, etc., i=1,…,N_s; N_s is amount of the stages of beneficiation TP); NC_{ij} – intelligence neurocontroller of OC_{ij}; V_{ij} is a vector of disturbing influences for OC_{ij}; Y_{ij} – a vector of output characteristics of OC_{ij}; U_{ij} is a vector of control influences (actions) of OC_{ij}; X_{ij} is a vector of informative parameters about the state of OC_{ij}; Y^s_{ij} is a vector of settings of output characteristics of OC_{ij}; TP^*_i is the complex of all local TP of the certain stage; V^*_i is a vector of main influences of disturbing of TP^*_i; Y^*_i is a vector of output characteristics of TP^*_i; X^*_i is a vector of information parameters about current stat of TP^* complex $_i$; Y^{*s}_i is a vector of tasks (settings) for output characteristics of TP^*_i; NE^*_i – neuroemulator (predictive mathematical model or predictor) for TP of the corresponding stage.

Three main control levels 1) of local regime parameters (ore and/or water consumption, pulp density, etc.); 2) quality indices (content of useful component,

output, exception, etc.); 3) complex of TP (fragmentation, classification, magnetic separation) are divided in the structure.

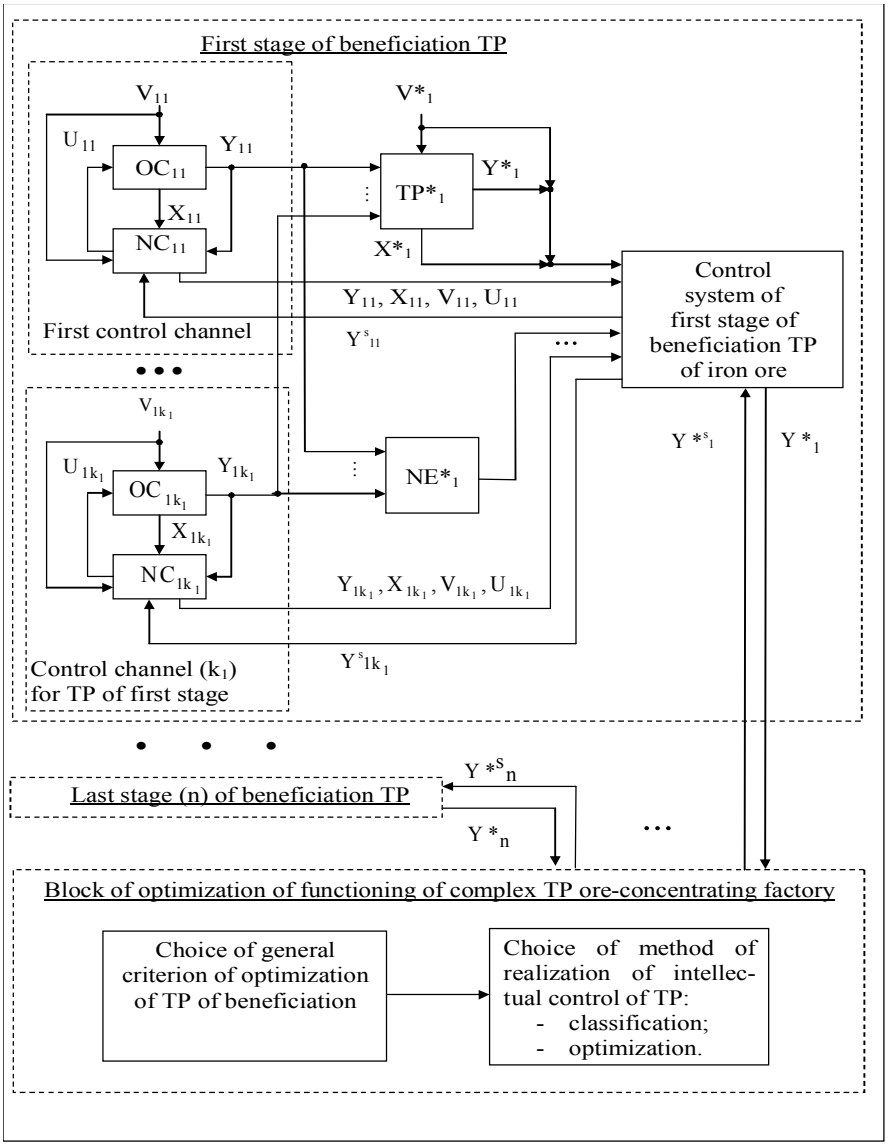


Fig. 2. The structure of combined multichannel ICS of TP of magnetite quartzites beneficiation (classification-optimal control)

So, for example, for a complex of TP of the first stage (supposing that for TP of fragmentation i=1, k_1=2): the first channel (OC_{11}) is the correlation of "ore-water"; the second channel (OC_{12}) is the mill productivity output (at unloading); V_{11}={coarseness of grading (averaged coarseness) of input product}; V_{12}={physical and chemical and

mechanical properties of ore}; Y_{11}, Y_{12}={ coarseness of grading (averaged coarseness) of industrial product, productivity after the industrial product, output of the prepared class}; U_{11}={mill water consumption }; U_{12}={ ore input productivity}; X_{11}={content of solid in the middle of the mill}; X_{12}={all regime indices of mill work}. Similarly the formalization for other TP of the first stage (classification, magnetic separation) is carried out. Then the resulting characteristics for a complex of TP (all stages) as a whole are formed as follows: $V^*_1=V_{11}\cup V_{12}$ ($\cup$ is the operation of logical combination of vectors); Y^*_1={quality of industrial product by quality of useful component, productivity on the output stage}; $X^*_1=X_{11}\cup X_{12}$.

The idea of the approach is in application of combined algorithm with combination of classification and optimal control approaches in order to ensure the acceleration decision-making process in multichannel ICS of TP of magnetite quartzites beneficiation. The main features of the implementation of such a system are as follows [1, 7].

The intellectual analysis of current state of control object is carried out constantly at the end of the next step of discrete time by the top level of the system on every stage of beneficiation in the block of optimization of beneficiation complex operation. The determining of settings (tasks) for the control systems of the corresponding stages (middle level) is carried out on the basis of a coherent analysis of indexes of all beneficiation stages. At the same time, in contrast to existing approaches, decision-making process (definition of the necessary settings) in the system (Fig. 2) can be occurred through intelligent classification (classification control) or global optimization (optimal control). Algorithms for the implementation of corresponding computational procedures will be given in the future.

On the middle level control of TP complex for separate stages is carried out. For this purpose the level is given the value of optimal settings from a top level and it determines a task (proves these settings) for the regulators of all local TP and their corresponding channels of control of every beneficiation stage. From the other side middle level systems collect primary information about the state of every channel (control actions, outputs, disturbing) from the subsystems of the bottom level, carry out its primary processing, prediction of values of input and output indexes of the stage using of neuroemulator (NE*i). Certain data are also passed on the top level for decision making and determination of optimal settings for the purpose of the coordinated control of all stages and complex of beneficiation TP as a whole.

The bottom level of the system controls separate local TP of each stage. For this purpose the level contains the number of control channels. Each channel has its own inverse neuroregulator that recreates the inverse dynamics of the process. The task of work of such regulator is maintenance of necessary value of settings, that is determined at the top level of the system and given from the corresponding control subsystem of the certain stage (id est. middle level). In turn, the bottom level subsystem passes information about the state of each channel (indexes of control influences, value of output and information signals, disturbance) to the middle level system at first and then to the top level.

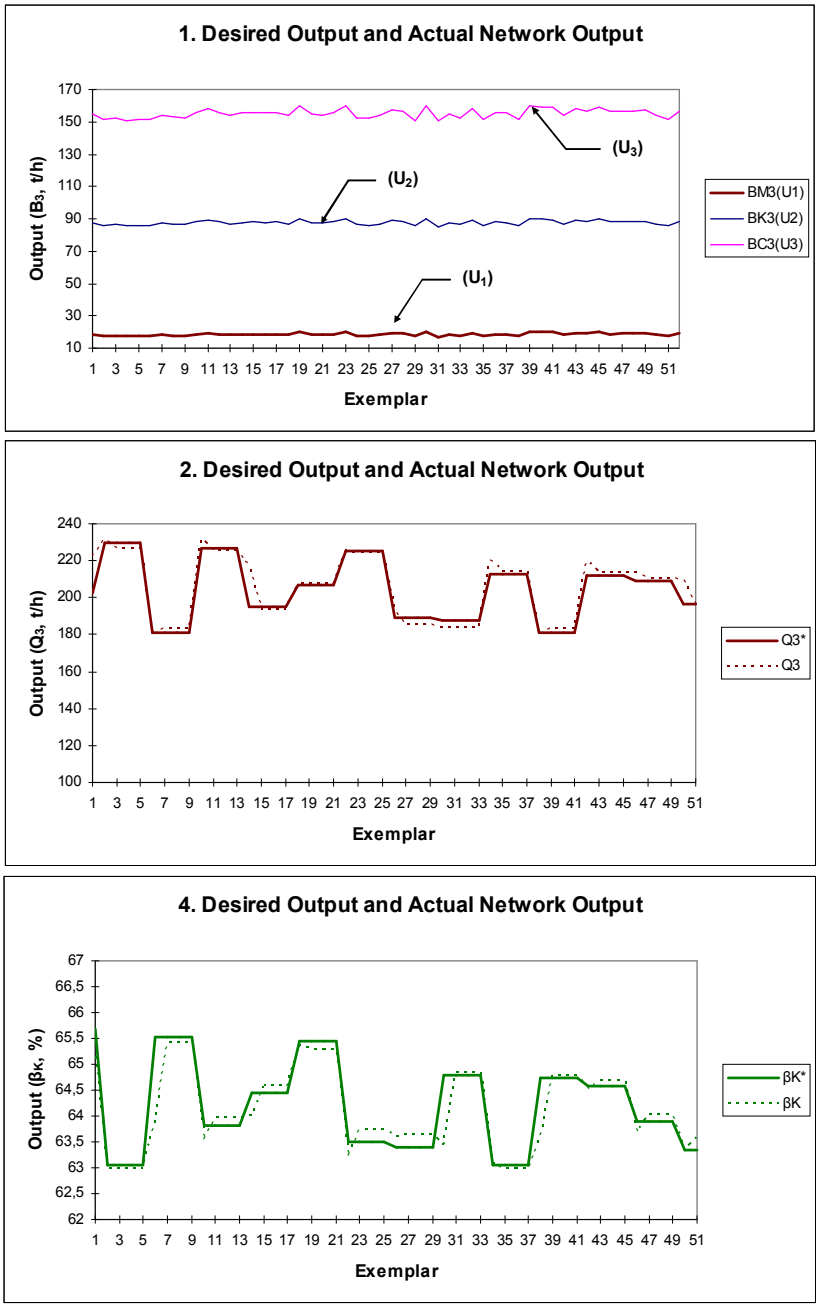


Fig. 3. Results of computer modeling of classification optimization process in the context of actual indicators of magnetite quartzites concentration

4 Conclusions

For the hierarchy scheme of ICS of beneficiation technological complex on the basis of combination of principles of neurocontrol, intelligence classification and global optimization contingency approach at forming limit cluster of certain "special technological situations", that allows to control TP automatically in real-time mode, determine and propose corresponding control influences has been offered.

The conducted researches, results of computer modeling (Fig. 3) and industrial tests [1, 5-7] proved that application of neural networks schemes on the basis of inverse models and neuroemulators as regulators of separate channels of beneficiation TP has a sufficient dynamics (reasonable time of settings exercise on condition of its presence), the possibility of the proper disturbance rejection at 10% level and operation on the conditions of nonlinear limitations (changes of controller parameters) on the basis of satiation principle. Thus, the task of this work is the verification of possibilities of classification strategy for reliable determination of optimal values of current parameters of TP (in the form of the relevant tasks or setting for controllers), that will provide stable work of local regulators in the above-mentioned terms.

References

1. Kupin, A.I.: Intellectual identification and controls in the conditions of processes of concentrating technology. The monograph. Kyiv: Korneychuk's Publishing house (2008)
2. Scheiner, B.J., Stanley, D.A., Karr, C.L. Emerging computer techniques in the minerals industry. Littleton, CO: Society for Mining, Metallurgy, and Exploration, Inc. (1993)
3. Wills, B.A. Automatic control in mineral processing . Mining Mag. No 3, pp. 316--320 (1987)
4. Maryuta, A.N., Kochura, E.V. Economic - mathematical methods of optimum control of the enterprises. Dnepropetrovsk: Science and education (2002)
5. Kupin, A.I. Neural identification of technological process of iron ore beneficiation. Proceedings of 4th IEEE Workshop on Intelligent Data Acquisition and Advanced Computing Systems Technology and Applications (IDAACS'2007). Dortmund, pp. 225--227 (2007)
6. Kupin, A.I. Research of properties of conditionality of task to optimization of processes of concentrating technology is on the basis of application of neural networks. Metallurgical and Mining Industry, No4, pp. 51--55 (2014)
7. Kupin, A.I. Application of neurocontrol principles and classification optimisation in conditions of sophisticated technological processes of beneficiation complexes. Metallurgical and Mining Industry, No6, pp. 16--24 (2014)
8. Scheiner, B.J., Stanley, D.A., Karr, C.L. Control of liquid level via learning classifier system. Proceedings of The Applications of Artificial Intelligence VII Conference. No1095, pp. 78-85 (1989)
9. Krasnopoyasovsky, A.S. Information synthesis of intellectual control systems. Sumy: Publishing house SumSU (2004)
10. Morkun, V.S., Tron V.V. Ore preparation multi-criteria energy-efficient automated control with considering the ecological and economic factors. Metallurgical and Mining Industry, No5, pp. 4--7 (2014)

A Composite Indicator of K-society Measurement

Kseniia Ilchenko[1,2], Ivan Pyshnograiev[1,2]

[1] World Data Center for Geoinformatics and Sustainable Development,
Peremohy av. 37, 03056 Kyiv, Ukraine
[2] National Technical University of Ukraine „Kyiv Polytechnic Institute“,
Peremohy av. 37, 03056 Kyiv, Ukraine

{ilchenko, pyshnograiev}@wdc.org.ua

Abstract. The development of K-society theories leads to necessity of finding an approach of measuring the progress of each country. The paper presents the composite model which based on OECD and UN methodology. The hierarchy model consists of three dimensions and 14 indicators and gives a possibility to calculate K-society Index for 87 countries. The analysis of the results presents country's current rating and dynamics. The data for Top-20 countries, the last twenty countries and North America are introduced in the paper. K-society Index for Ukraine is described in details. The future state's strategy can be based on K-society measurement.

Keywords: MathematicalModeling, Knowledge, Methodology, DecisionSupport.

1 Introduction

The fundamental concept of sustainable development requires the review of the classic studies about the world. Knowledge as a higher value of informational process forces the progress in sustainability. Also, the knowledge is one of the factors of production in modern economy. That is why, the theory about knowledge-based economy and society becomes wide shared among scientists. For example, knowledge society is described as sustainability concept by N. Afgan and M. Carvalho [1]. M. Kulin studies learning and knowledge influence as a factor of global competitiveness [2], the impact of knowledge for society is the main focus of G. Bohme and N. Stehr research [3]. Thus, the theoretical aspects of Knowledge society are well-studied.
At the same time, the questions about applied evaluation of knowledge in a country, comparison of different countries and knowledge dynamics research are still open.
Taking into account the complex character of knowledge, it can be presented as the set of indicators which are gathered in a hierarchy model.
Therefore, the main idea of the research is to draw out a composite indicator for measurement knowledge as a sophisticated category with a purpose of country development analysis.

2 K-society as a New Mode of the World Developing

Classic economic theory presents three factors of production that are used in a production process, which leads to finished goods. These three basic resources are land, labor and capital. Nowadays this fundamental approach was divided into several complex theories that include additional factors of production, for example, technological progress, human capital and social capital. Basically, those resources can be aggregated into one category – knowledge. More than this, knowledge and information become the most significant factors of production and form the basis for new technological mode.

Knowledge society (K-society) is widespread concept, but scientists still investigate its nature [4]. The mass production of knowledge changes the economy in global world in quite short terms. However, this process is dissimilar in different countries.

The research of K-society is undertaken by all developed countries for more than 40 years but there are still a lot of controversial question. First of all there is no agreement about terminology. Such terms as “K-society”, “Informational society”, “Technogeneous society” serve the purpose of science communication in this topic. The term “K-society” was used by M. Zgurovsky to mean a where institutions and organizations give possibilities to people and information to develop without any barriers and open opportunities for mass production and mass usage of all kind of knowledge in global scales. Therefore, the development of technologies is an important part of K-society, but not the main purpose. Thus, the term “Technogeneous society” doesn’t describe these processes in full measure.

The question about links between K-society and information society is more complicated. The first one is based on definition of knowledge, the second one uses information as a basic category. The development of new computing technologies has not influenced to significancy of common paradigm, but the possibility to get, safe, analyze and transfer knowledge was changed cardinally. That led to increasing velocity of information circulation. Moreover, it is difficult to divide information and knowledge. But in the purpose of this research it is assumed that knowledge includes information, and it is a product of information processing.

According to theoretical research the concept of K-society is ambiguous. On the one hand it is a philosophic theory, which has no practical meaning, on the other hand it is the set of instruments and methods for providing sustainable development of modern society [5]. In accordance to the second opinion K-society proclaims the active usage of knowledge, which is the main asset.

The main accent is education, which forms a human capital and guarantees the access to information. But the measurement of educational level cannot give a complete picture of knowledge in society. Therefore, K-society must be formalized more manifold system of indicators. Likely, such system includes the description of current situation in economy, perspectives and information transactions. It is obvious that the development of model for describing K-society is a nontrivial issue.

3 Methodology

According to the UNO methodology, the index of K-society should be based on three dimensions: Assets, Advancement and Foresightedness [6]. The first one describes the level of education, especially, among young people, and the development of information streams. These two main directions include such indicators as: expected schooling, proportion of young people, the diffusion of newspapers, the Internet, main phone lines and cellular phones. The second dimension represents human and informational resources, which are indicated by public health expenditure, research and development expenditure, military expenditure, pupil/teacher ratios in primary education, and a proxy of the "freedom from corruption" indicator. The last dimension shows the external influence on K-society dynamics in the state. This dimension consists of low child mortality rates, equality in income distribution (GINI Index), protected areas as percentage of a country's surface, and CO2 emissions per capita indicators. This approach was officially accepted for approximately 45 countries in 2005.

Taking into account the existent basic specification of the main categories, it becomes possible to continue this research in terms of current informational mode. Thus, new hierarchical model for K-society measurement should be built.

Therefore, it is necessary to clarify the approaches for drawing out this model. The OECD presented methodology and user guide on constructing composite indicators [7]. According to this, there are several obligatory steps in models' creation.

Firstly, the full understanding of processes that can have influence on K-society needs to be represented in theoretical framework. This step concludes with the number of selection criteria. As referred to listed above, the framework is based on UN model.

Secondly, the very important step is data selection. It includes the availability and quality data checking. In addition, the question about strengths and weaknesses of indicators must be resolved. Not least important is to find the reputable source for each data set. Theoretically, all data must be provided by international world-known organizations.

In view of these two steps the UN approach has some disadvantages that are caused by following reasons. On the one hand, last ten years have brought significant changes in informational development. As a result of this process some of indicators lost their relevance. On the other hand, not all data sets are still gathered by authoritative organizations. That is why the original model needs revision and modernization.

Thirdly, the modeling needs complete data sets. Thus, the problem of empty cells that usually appears after the data selection requires imputation of missing data. The various kinds of methods for working with complex models are established in World Data Center for Geoinformatics and Sustainable Development [8]. Therefore, the recommendation for this case is to augment the empties by previous period information.

The step includes multivariate analyses. This phase gives the possibility to double check the starting hypothesis about the set of indicators. The significance of sampling should be checked. Other important question is to evaluate relations between indicators. That is why the elements of principal components analysis and cluster

analysis influence the final decision about model structure. This step identifies statistically similar indicators. Thus, the additional explanation of internal relations or model's rebuilding can be required. As a result of this issue the model can be amplified by additional explanation.

Taking into account the miscellaneous nature of indicators the next step is normalization. There are more than ten typical approaches to its implementation. It is necessary to underline that there is no goal to make the estimation more complex. For this reason, the standardization is the optimal variant for this step. The formula of this type of normalization is as below:

$$Value_{norm} = (Value - Value_{min})/(Value_{max} - Value_{min}) . \quad (1)$$

In case when it is necessary to represent the inverse coupling this formula converts to:

$$Value_{norm} = 1 - (Value - Value_{min})/(Value_{max} - Value_{min}) . \quad (2)$$

As a result, all indicators values lie in interval from 0 to 1.

To express the theoretical framework and relations underlined at the previous stages, the sixth step includes finding out the way of indicators aggregation and their weights establishment. For instance, the model's hierarchy is constructed.

Each dimensions' index consists of several indicators and can be presented as the average value of its components. In the same manner K-society Index equals to the all dimensions indices aggregation.

At the next step uncertainty and sensitivity analysis emphasize the reasons of the differences between results of using variety of aggregation, imputation and normalization methods. This step identifies all possible sources of uncertainty and determines what sources have more influence to the overall score.

Eighthly, detecting dominant and critical indicators for objects or their groups provides the information about the levels of influence for the assessed system. It is also very important for policy making problems.

Then, for modeling results validation developed index is compared to others that describe the phenomenon of similar nature. The comparison base consists of well-known indices that authoritative organizations and institutions provide. Thus, two indices were chosen for purposes of final analysis: Fragile State Index [9] and Index of Economic Freedom [10].

Finally, the last step is to present the results in a clear and accurate manner. That is why visualization is the part of this algorithm. It is necessary to choose the correct tools that provide total understanding of the obtained results. Thus, the final step of modeling becomes the element of a decision making support system.

Taking into consideration the UN approach and OECD methodology the new model was drawn out. The indicators, data providers and data sources are presented in Table 1.

Table 1. List of indicators

Indicator	Institution	Source	Type of influence
School life expectancy	UNESCO Institute for Statistics	http://www.uis.unesco.org	Positive
School enrollment, secondary (% net)	World Bank	http://data.worldbank.org/indicator/SE.SEC.NENR	Positive
Internet subscriptions per 100 inhabitants	ITU	http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx	Positive
Main phone subscriptions per 100 inhabitants	ITU	http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx	Positive
Cellular subscriptions per 100 inhabitants	ITU	http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx	Positive
Gov't Health Expenditures (% of total gov't exp)	World Health Organization	http://apps.who.int/gho/data/?theme=main	Positive
R&D expenditure as % of GDP	UNESCO Institute for Statistics	http://www.uis.unesco.org	Positive
Military expenditures (% of GDP)	SIPRI	http://www.sipri.org/	Negative
Pupils per teacher in primary school	World Bank	http://data.worldbank.org/indicator/SE.PRM.ENRL.TC.ZS	Negative
Corruption perception	Transparency International	http://www.transparency.org/research/cpi/overview	Positive
Child mortality (children under 5 years per 1000 births)	World Bank	http://data.worldbank.org/indicator/SH.DYN.MORT	Negative
Gini Index	World Bank	http://data.worldbank.org/indicator/SI.POV.GINI	Negative
Terrestrial and marine protected areas (% of total territorial area)	World Bank	http://data.worldbank.org/indicator/ER.PTD.TOTL.ZS	Positive
CO2 emissions (metric tons per capita)	World Bank	http://data.worldbank.org/indicator/EN.ATM.CO2E.PC	Negative

Data for 87 countries were gathered and complemented in the process of model development. Thus, the results of estimations are described in the next paragraph.

4 Results

According to the algorithm each of the dimensions were counted based on their components. It is necessary to mention that it gives the possibility to measure Assets, Advancements and Foresightedness as separate indices. Such evaluation brings an opportunity to additional comparison of countries in terms of the dimensions. But in accordance with the main purpose of the research the K-society Index has to be measured. That is why the procedure of linear convolution is implemented twice.

Collected data give a possibility to provide the calculations for period from 2008 to 2013.

The results for 2013 year show that the Top 10 countries for K-society Index consists of Switzerland, Denmark, Netherlands, Sweden, Slovenia, France, Austria, New Zealand, Japan and Finland. The values for the final index and three dimensions are presented in Table 2.

Table 2. Top 10 countries by K-society Index 2013

	The Assets Index	The Advancement Index	The Foresightedness Index	KS Index	Rank
Switzerland	0,801	0,827	0,780	0,803	1
Denmark	0,758	0,794	0,785	0,779	2
Netherlands	0,764	0,757	0,789	0,770	3
Sweden	0,722	0,809	0,766	0,766	4
Slovenia	0,670	0,642	0,949	0,754	5
France	0,789	0,636	0,792	0,739	6
Austria	0,703	0,749	0,765	0,739	7
New Zealand	0,744	0,737	0,711	0,731	8
Japan	0,719	0,777	0,687	0,728	9
Finland	0,692	0,773	0,722	0,729	10

The analysis of representatives shows that Top 10 involves high-developed countries with sustainable economic, ecological and social conditions. The variance between the first and the last states from the list described above equals to 0,074. Moreover, the gap between top possible value of the index, and the value for Switzerland is 0,197.

The last 10 countries of the ranking for 2013 year are presented in following table (Table 3).

The last one, Nigeria, has a high level of Fragile States Index, which is caused by alert meaning of such indicators as Demographic Pressure, Group Grievance, Uneven Economic Development, State Legitimacy, Public Services, etc. Even more, the conflict barometer, which is counted by HIIK [11], shows that this country has the value 5. That means the existence of the war in Nigeria.

According to the same sources Pakistan is under the inter-ethnic violence and conflict with India that were classified as limited war and violent crisis. Also the problems with Demographic Pressure, Refugees, Group Grievance, State Legitimacy

Human Rights, Security Apparatus, etc. exist in the state. Moreover, the situation, described by Fragile States Index, is even worse than in Nigeria.

Table 3. Last 10 countries by K-society Index 2013

	The Assets Index	The Advancement Index	The Foresightedness Index	KS Index	Rank
Paraguay	0,280	0,359	0,569	0,402	78
Senegal	0,145	0,432	0,633	0,403	79
India	0,278	0,327	0,589	0,398	80
Madagascar	0,133	0,334	0,543	0,337	81
Gambia	0,145	0,359	0,455	0,320	82
Kenya	0,206	0,249	0,492	0,315	83
Ethiopia	0,052	0,284	0,637	0,324	84
Mozambique	0,183	0,260	0,494	0,312	85
Pakistan	0,107	0,182	0,578	0,289	86
Nigeria	0,074	0,304	0,433	0,270	87

The next one is Mozambique. In accordance to the Fund for Peace methodology the state's current pressure assessment is "Very High Warning". The more dangerous indicators are: Demographic Pressure, Uneven Economic Development, Economy and Public Services.

Ethiopia is in a group of countries, which have "alert" status. The greatest problems of Ethiopia are Social and Economic Fields, External Intervention and Factionalized Elites. Such tendency has been continuing since 2009.

Kenya has a limited war, which is connected with inter-ethnic violence. In addition this state is 18 from 178 countries in Fragile States Index. The problems with Political and Military, Social and Economic fields lead to high negative rating.

The next country is Gambia. It has growing tendency from stable to very high warning assessment in Fragile States Index.

India is the neighboring country for Pakistan. Thus, problems with conflicts, which were described above, also concern India. Furthermore, India has to worry about Demographic Pressure, Group Grievance, Uneven Economic Development and Security Apparatus. The less number of problems gives India higher value of K-society Index. The fact of common knowledge is that India tries to develop IT sphere. But it seems that it is not enough for building K-society.

Senegal has stable, very high warning assessment since 2006. The long-term tendencies show that the situation in the country becomes more and more dangerous. Madagascar is near Senegal in rating and the common tendencies almost the same, except the reduction of Group Grievance and Refugees. Such situation has been occurred since 2008. However, Paraguay is the only country from the bottom part of the rating that has been increasing in Fragile States Index in terms of improving situation.

This analysis shows that K-society Index reflects much more information than IT or science alone. It correlates with current political and economic situation in the country. Moreover, it is impossible to build K-society in unsustainable environment.

It is essential to discover the relations between K-society Index and other well-known indices. Fig. 1 shows the correlation between Fragile States Index and K-society Index.

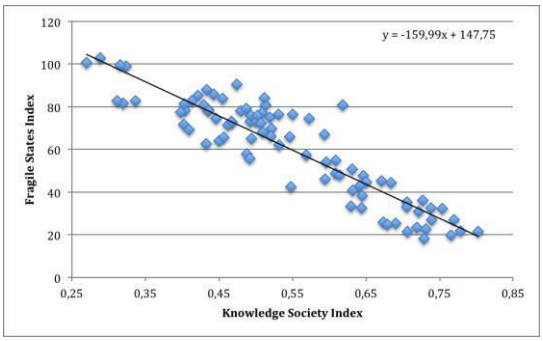


Fig. 1. Correlation between Fragile States Index and K-society Index

It describes high linear relation between indices. Thus, it is an additional proof of state instability influence to knowledge establishment.

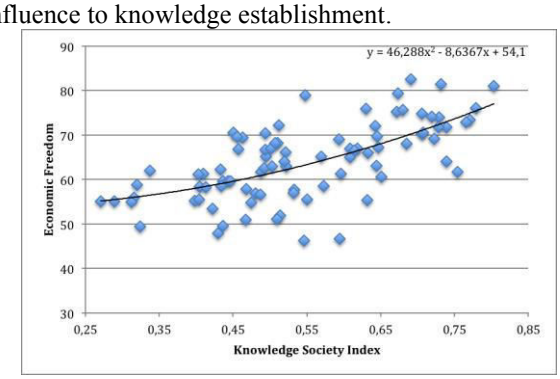


Fig. 2. Correlation between Economic Freedom and K-society Index

Probably, more interesting results were obtained from K-society Index and Economic Freedom relations. Fig. 2 shows that the economic component is not fundamental for processes in K-society. The truth is that economy is rather important.

The results of the research show that K-society can be unequal in neighboring countries. Also there is no dependence between the leading positions in the world and absolute success in K-society creation. For instance, the comparison of Mexico, USA and Canada is a good illustration of mentioned above thesis (Fig. 3).

The graph illustrates the North America countries' values. The first place has Canada. The USA shows almost the same tendency but with lower score. Both countries have falling K-society Index tendency in 2012-2013. It is noteworthy that Mexico's tendency corresponds to others but the values of index are much lower on all period of research.

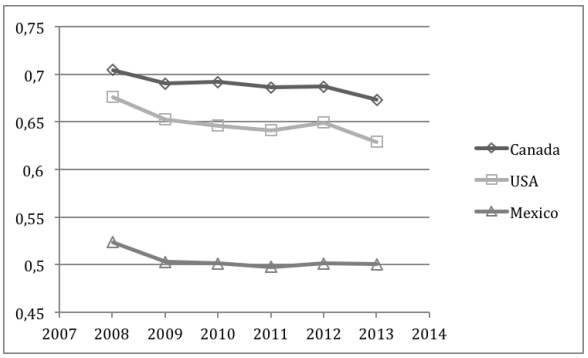


Fig. 3. K-society Index for Mexico, USA and Canada

The challenging issue is to find out Ukraine's situation with K-society development. Ukraine had good infrastructure, science and educational bases but it is necessary to clarify it is still competitive or not in the international area.

The first step in this direction is to compare Ukraine with neighboring countries. Taking into account that all neighbors are from post-Soviet area, this sample is congeneric. Thus, the results in the index form should describe the Ukrainian success in K-society development. In addition, the qualitative information about neighbors gives a possibility to verify calculations. The existence data let to find values of index for Poland, Russia, Moldova and Hungary. The dynamics of K-society Index for these countries and Ukraine is introduced on one graph. This approach allows demonstrating the differences obviously (Fig. 4).

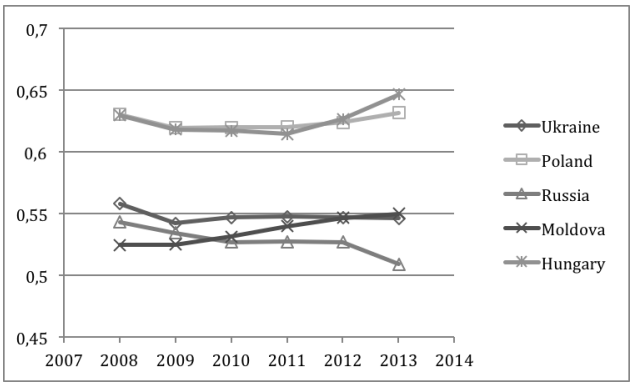


Fig. 4. K-society Index for Ukraine's neighbors

Firstly, it is necessary to mention that Russia confirms the significant fall of index' values in 2012-2013, that USA and Canada showed. Secondly, two countries, Hungary and Poland, have almost equal dynamics of index' values. Ukraine shown higher estimations than Moldova and Russia in 2008 and outstripped those countries until 2012. The situation was changed in 2013 when Ukraine got lower position than Moldova. In general, Ukraine takes the 40th place from 87 countries in 2013. Its value

of K-society Index equals to 0,546. It is to be recalled that the value for Switzerland is 0,803.

It is useful to discover the components of index for Ukraine to define the weak part of it. Fig. 5 illustrates the Assets, Advancement, Foresightedness and K-society Indices' dynamics from 2008 to 2013.

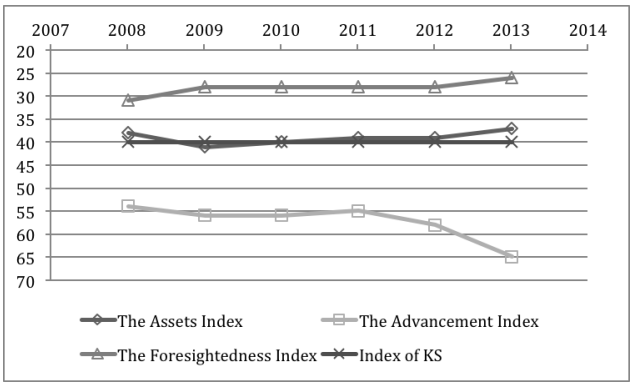


Fig. 5. Ukraine's values of K-society Index and its components

In the purpose of this analysis all components are described by places rating. This gives the opportunity to show relative measures and ranking. The Advancement dimension shows the worst values in all period. Thus, let's consider from what indicators this dimension consists of. Obviously, Ukraine has a great problem with freedom from corruption indicator. In addition, research and development expenditures, pupil/teacher ratios in primary education and public health expenditure are lower than generally accepted (for example, in Europe) norms. This issue can be an opportunity to significant development of K-society in future. Accordingly, these fields need to be modernized and get all possible funding for improving the situation. Thuswise, this analysis shows the preconditions of strategic planning and decision making in Ukraine in case it is necessary to reach the leading countries. The last hypothesis is based on the fact that the leaders in K-society Index are the most developed countries.

5 Conclusions

In paper it was shown that K-society is a probable next mode of economy development that leads to changes in institutional and organization structure inside each country and over the world.

K-society is a complex category, which can be considered as a strategy goal for country. Therefore, it needs to be measured in quantitate form. The analysis of existence approaches shows that it is possible to use OECD methodologies for creating composite indices and UN methodology for K-society Index. The

improvement and combination those two sources give the base for model of K-society Index.

The K-society Index was drawn out as a combination of three dimensions and 14 indicators. The values of index were calculated for 87 countries that provide all necessary information.

The analysis of results shows that there is no direct dependence between K-society development and the country leadership in the world.

The situation for Ukraine was analyzed deeply. Firstly, Ukraine has lower meaning of index than it's neighbor countries Moldova, Poland, Hungary. Secondly, the less developed dimension is "Advancement". Thus, the strategy of its extension must be provided.

Some common tendencies were found for all countries. The index decreased rapidly its value in 2008. The values of index have high correlation with Fragile State Index and Economic Freedom.

References

1. Naim Hamdija Afgan, Maria G. Carvalho. The Knowledge Society: A Sustainability Paradigm. Cadmus. Volume 1. Issue 1 (2010)
2. Michael Kuhn, Massimo Tomassini, P. R. J. Simons. Knowledge Based Economy: Knowledge and Learning in European Educational Research (2006)
3. Gernot Bohme, Nico Stehr. The Knowledge Society: The Growing Impact of Scientific Knowledge on Social Relations. Springer Science & Business Media (1986)
4. UNESCO World Report: Towards Knowledge Societies. UNESCO, France (2005)
5. Understanding Knowledge Societies in Twenty Questions and Answers with the Index of Knowledge Societies. New York: UNPAN (2005)
6. Understanding Knowledge Societies. Department of Economic and Social Affairs, United Nations, New York (2005)
7. Handbook on Constructing Composite Indicators: Methodology and User Guide. Organization for Economic Co-Operation and Development, France (2008)
8. World Data Center for Geoinformatics and Sustainable Development, http://wdc.org.ua/
9. Fund for Peace: Fragile State Index, http://ffp.statesindex.org/
10. Terry Miller, Anthony B. Kim, Kim R. Holmes: Highlights of the 2014 Index of Economic Freedom: Promoting Economic Opportunity and Prosperity. The Heritage Foundation, New York (2014)
11. Conflict Barometer 2013. Heidelberg Institute for International Conflict Research, Germany, №22 (2013)

Implementing Manufacturing as a Service:
A Pull-Driven Agent-Based Manufacturing Grid

Leo van Moergestel[1], Erik Puik[1], Daniël Telgen[1], and John-Jules Meyer[2]

[1] HU Utrecht University of Applied Sciences, Utrecht, the Netherlands
{leo.vanmoergestel, erik.puik, daniel.telgen}@hu.nl
[2] Utrecht University, Utrecht, the Netherlands
J.J.C.Meyer@uu.nl

Abstract. User requirements and low-cost small quantity production are new challenges for the modern manufacturing industry. This means that small batch sizes or even the manufacturing of one single product should be affordable. To make such a system cost-effective it should be capable to use the available production resources for many different products in parallel. This paper gives a description of the requirements and architecture of an end-user driven production system. The end-user communicates with the production system by a web interface, so this manufacturing system can be characterized in terms of cloud computing as the implementation of manufacturing as a service, abbreviated to MaaS.

Keywords: agile manufacturing, agent technology, MaaS
Key Terms Industry, Infrastructure, Machine Intelligence.

1 Introduction

At the HU Utrecht University of Applied Sciences, an agile manufacturing system has been developed that is capable of so-called multiparallel production of small batches or even one single product. The need for such a manufacturing system comes from the fact that nowadays the demand for custom end-user specified products is increasing. Internet is offering a method to involve the end-user directly into the production. Also the possibilities of additive manufacturing by using 3D printers offers new ways to set up a manufacturing infrastructure with the focus on the manufacturing of small quantities.

This paper will focus on the interface to connect the end user to the production process. Before going into detail, the manufacturing system itself will first globally be described.

In the next section details about the basic design considerations are given. Because the implementation is based on agent technology, a short description of what an agent is, will be given. The architecture and the connection with the end-user will be the treated next. Finally, the results, related work, discussion and a conclusion will end the paper.

2 Global description of the manufacturing system

Every product to be made starts its life as a software entity, that contains the information what should be done to make the product. This software entity is a so-called software agent.

2.1 Agents

A common definition of an agent given by Wooldridge and Jennings [13] is:

Definition (agent). An agent is an encapsulated computer system or computer program that is situated in some environment and that is capable of flexible, autonomous action in that environment in order to meet its design objectives or goals.

The manufacturing system that has been designed is based on a group of cooperating agents. A system with two or more agents is called a multiagent system (MAS). In our design the following properties of agents are important:

- goal: an agent is designed to reach a goal. If reaching the goal is complex, subgoals can be defined as states to be reached to finaly come to the end-goal.
- action: an action is what the agent can do.
- plan: to reach a goal or subgoal the agent builds or receives a plan. Normally a plan consists of a list of actions to reach a goal or subgoal within a certain role.
- role: agents can have different roles. In a multiagent system these roles play an important part in the way agents cooperate.
- behaviour: closely related to the role is the behaviour. This the set of actions that an agent will perform in a certain role.
- belief: a belief is what the agent expects to be the case in the environment.

In multiagent technology other aspects can also be important, but the properties mentioned here are specific for the manufacturing system presented in this paper. The main reason for choosing agent technology is that it offers a natural decomposition of responsibilities and tasks to be completed in this complex manufacturing system. It also means that if one agent fails, other agents can continue to fulfil their own goals and even take over actions or tasks from the failing agent.

2.2 The manufacturing grid

The infrastructure of the manufacturing system consists of cheap reconfigurable production machines that we will call equiplets. These equiplets are capable to perform one or more production steps. The set of steps an equiplet can perform depends on its front-end. An equiplet can be reconfigured by changing its front-end. The equiplets are placed in a grid arrangement. In conventional mass production, a line arrangement is used because for all products the same sequence of production steps should be followed. However in our case every product can have a different path along the equiplets, so a grid arrangement is more natural offering multiple mostly shorter paths in case of an arbitrary sequence of equiplets to be visited.

2.3 Agent-based manufacturing

The equiplet-based manufacturing description will have its focus on the MAS where the equiplet agent is the representative of the equiplet. An equiplet agent will publish its capabilities. This means it will announce its production steps. It will wait for products to arrive to actually perform the production steps.

The product agent has several roles. It starts with planning the path along the equiplets for the production. Next, it will schedule the production. After successfully scheduling, it will guide the product along the equiplets. At every equiplet it will instruct the equiplet agent what step or steps to perform. It will log the results of a production step and also update a globally shared knowledge base that can be consulted by other product agents to check the reliability of a certain equiplet for a certain step with certain parameters. Having the responsibility for the manufacturing of a product, the product agent is also the entity that should recover from errors during manufacturing. If there is a failure on a certain equiplet, depending on the type of failure (recoverable or severe) the product agent will try to plan the required step on an alternative equiplet for the same reason as why one would not prefer to hire a plumber who previously made mistakes resulting in a flood. By putting the information about the failure (step type and parameters) in a shared knowledge base, the product agents will learn as a group about the reliability of the equiplets for certain steps.

When the product is finished, the product agent can also have a role in other parts of the life cycle of a product, being a software entity that knows a lot about the product and the actual production. To achieve these roles, the agent could be embedded in the product itself, but being accessible in cyberspace is also a possibility.

3 System architecture

In this section a description of the system architecture as well as the constraints on our type of production will be presented.

In figure 1 the layered software architecture is given. Only one product agent and one equiplet agent is depicted and the modules in the lower layer of the equiplet depend on the front-end that has been connected to the equiplet. In this case an equiplet with the pick and place capabilities and vision modules is used in this example. For the MAS layer Jade [1] was used as a platform. Jade is a widely accepted Java-based multiagent environment. The inter-agent communication is implemented by using blackboards. A blackboard is a software entity where agents can publish information that will be available to other agents.

The software for the equiplet is based on ROS. ROS is an acronym for Robot Operating System [10]. ROS is not really an operating system but it is middleware specially designed for robot control and it runs on Linux. In ROS a process is called a node. These nodes can communicate by a publish and subscribe mechanism. In ROS this communication mechanism is called a topic. This platform has been chosen for the following reasons:

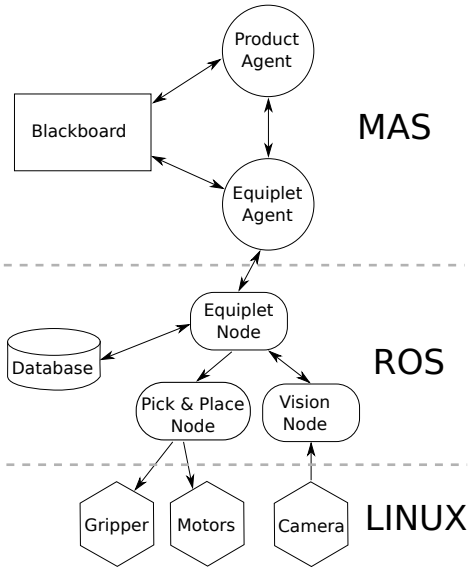


Fig. 1. Layered architecture

- Open source, so easy to adapt, compliant with a lot of open source tools.
- Wide support by an active community.
- Huge amount of modules already available.
- Nodes that are parts of ROS can live on several different platforms, assumed that a TCP/IP connection is available.

At the lowest layer in figure 1 is a Linux platform running modules that communicate with the underlying hardware. Linux is a stable, portable and versatile platform. In the next section we will take a closer look at the implementation of this architecture in combination with a web interface.

Our production model is based on trays that will carry the product to be built. These trays are transparent boxes, so equiplets with a camera can inspect them both from the top and the bottom. In the latter case the workplace of an equiplet should also be transparent, which is the case for the equiplets built so far. The trays are marked with a unique QR-code. During the first production steps the trays are filled with all the components required to make the product. This way a kind of construction box is generated. This means that for all steps to come, the components are available. This is a big advantage over a situation where logistic streams of components within the grid should be taken care of. The disadvantage is that parallel production of sub-parts in complex production paths is not possible. However for the proof of concept this is not a big problem and solutions can be found where the sub-parts are first manufactured in parallel and added to the construction box. Of course within our conceptual model other production models could be used, but the examples given here are based on this model.

4 Connecting the end-user

To use the manufacturing grid, a webserver has been added to allow end-users to construct products to be made by the grid. This is why it can not happen that a product is requested that does not fit within the capabilities of the manufacturing grid, because the grid itself is offering the webinterface for designing the product. If a product can be made using the webinterface, the grid will be capable to make it. This web interface will be called WIMP as an acronym for Web Interface Managing Production. The addition of a web interface as shown in figure 2

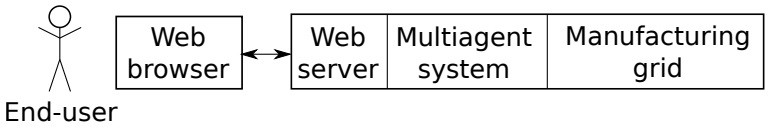


Fig. 2. Combination with webinterface

fits neatly in the concept of agile and lean manufacturing [11], where the end-user plays a prominent role in the production itself. The end-user specifies the product that will be tailor-made to his or her requirements. This pull-driven type of manufacturing will not lead to overproduction and waste of material.

The architecture of the software of the manufacturing system is depicted in figure 3. In this figure blackboards are abbreviated by BB. A web server

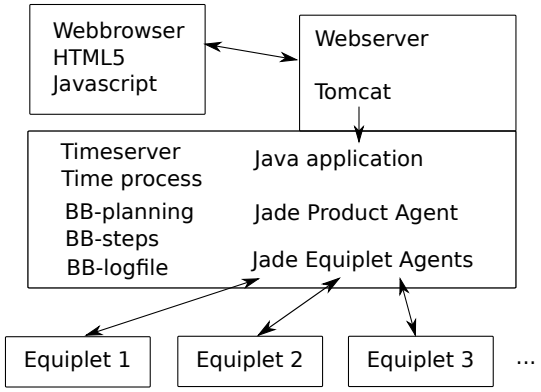


Fig. 3. Combination with webinterface

publishes a website where a customer can design his product. This could be a new product if the steps to produce it are within the capabilities of the equiplets in the grid. The webserver will be responsible to offer only those production step possibilities that are present in the grid. By pushing a submit button, a server-side program will create and activate a product agent. This agent will start to

plan the production path and communicate with the available equiplet agents to create the product. A more technical picture showing the distributed nature of the system is given in figure 4. The numbered components in figure 4 are:

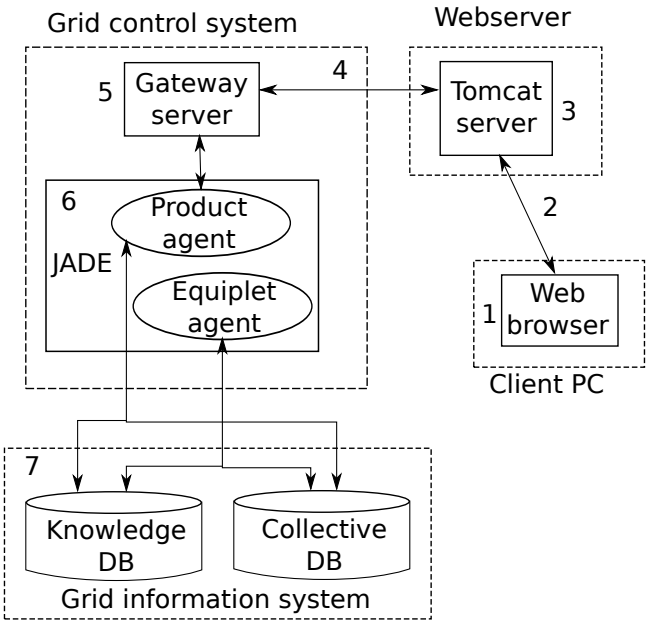


Fig. 4. Different platforms and their relations

1. The client PC as used by end-user. The end user can use any HTML-5 enabled browser.
2. Connection to the Tomcat server is established via a web socket.
3. The Tomcat server on which the website is hosted. The server can be placed on the grid server, but it can also be located somewhere else.
4. A connection between the gateway server and the Tomcat server is made through a (Java) socket.
5. The gateway server is responsible for spawning a product agent in the jade container. The gateway server acts as a *gateway* to the outside world, implemented to be able to spawn agents.
6. The Jade container of the grid contains all agents. Agents can communicate with the Tomcat server as will be explained in more detail further on in this paper.
7. The grid information system is a server where the databases and blackboards reside. These are the systems where shared and individual knowledge will be stored.

Agents have to be able to report back to the user. In order to do so, a software solution was implemented to allow them to send information over a socket. In

order to keep the connection alive, a heart-beat system has been developed. This is not shown in detail in figure 4, but the realisation will be described in the next sections.

4.1 Communications with the web interface/Tomcat server

Once a product agent is created through the web interface, the agent will create a socket behaviour. This socket behaviour is the way for a product agent to communicate with the server and thus to the web interface. To check whether or not the server is still alive and reachable a *heart* message is sent. If this message is not answered with a *beat* message it is assumed that the server is down. This is how the socket behaviour is used and implemented: The socket behaviour is used for the communication with the web interface and extends the Jade Waker behaviour which means it will become active after a certain amount of time. At the time of writing the wake up period for the socket behaviour is set at 5 seconds. This means that every 5 seconds the socket behaviour will become active and check if it is connected to the WIMP server. If it is connected it will check if there are data in the buffer; if any it will process the data. If the buffer is empty or if all data is processed the socket behaviour will go idle and will become active once the Waker behaviour is fired again after 5 seconds. The socket behaviour can also be used to write messages to the WIMP server even if the socket behaviour is not active, this is because it will be executed within the action method of another behaviour.

The heartbeat behaviour was created to eliminate a problem we were having with the socket behaviour. The problem encountered was the socket behaviour being unable to see if the socket connection is still alive, if it is not closed properly. The socket behaviour will only know if the connection is closed when either the client closed it properly or when the socket behaviour is trying to write on the socket when it is closed. Because we can receive commands from the WIMP server, we need to be sure the connection is active. If the connection is closed, but the socket behaviour is not aware of this, that would mean that the socket behaviour simply cannot receive messages from the WIMP server. And since the socket behaviour does not know the socket is closed, it will not try to reconnect. The heartbeat behaviour sends a *heart* message every 5 seconds and sets a timeout timer for 15 seconds. After sending a heart message the heartbeat behaviour expects a response within 15 seconds from the WIMP server. The response should be a *beat*. If it does not receive a response message within 15 seconds it will report to the socket behaviour that the connection is no longer active and will tell the socket behaviour to reconnect. If it is not possible to reconnect immediately, the socket behaviour will try to reconnect every time it becomes active.

4.2 WIMP capabilities

At the client side a web-browser receives a web-page in HTML5 format with embedded JavaScript and will display a graphical environment where a product

can be designed. This is the user interface of what has been called the WIMP. At this moment 4 typical product design web interfaces are implemented in WIMP:

1. Pick and place: 2D ball in cradle placement.
2. Paint pixels: pixel-based picture.
3. Pick, place and stack: simple 3D design.
4. Inspection of 3D printing object in STL-format.

A simple example of the pick and place interface is shown in a screen-shot in figure 5. A case with compartments of a certain dimension specified by the user

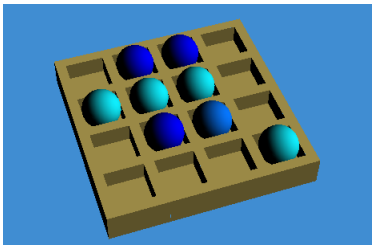


Fig. 5. Case with coloured balls in the webbrowser

is to be filled with coloured balls. The end-user selects a ball of a certain colour and moves the ball to an empty compartment.

An example of a sceenshot of the paint design interface is given in figure 6. On a canvas, a pixel-based painting using a combination of several colours can be made.

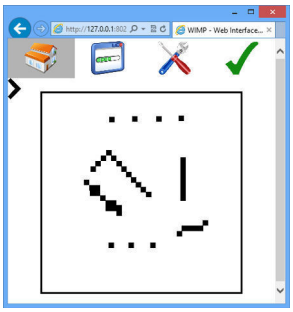


Fig. 6. A simple paint example

The WIMP software is also capable to build three-dimensional structures. It has some built-in intelligence. For example if a user wants to add a part at a place where adhesive is needed to keep it in place, it will warn the user

if he / she did not select the adhesive option for the placement of this part. This part of WIMP is only a basic implementation and in future development

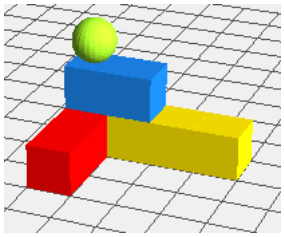


Fig. 7. A 3D structure

all kinds of special provisions should be added. For example when gluing two objects together several points of special interest arise. First of all the location of the objects you want to glue is very important. If the object is glued onto an existing structure it is possible that the existing structure will tip over. The structure must be stable enough and strong enough to support the new object. To determine if those conditions are met you have to know the material of the current structure, how much it weighs, and several other factors. Another important aspect of gluing objects is the type of adhesive. Not all materials can be glued together and not all types of adhesive can be used in combination with all materials. During manufacturing the objects that will be glued must be held together. This must be done until the adhesive is dry. Some adhesive types need heat to function properly, other types can be hardened by using UV-light.

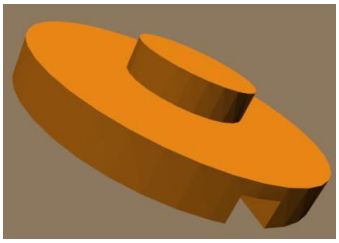


Fig. 8. View of an STL-image

At the client side a product is described by JSON. JSON, or JavaScript Simple Object Notation is a popular alternative to XML. XML was the de-facto standard before the existence of JSON. Until HTML 5, you needed to include libraries to encode and decode JSON objects. Now, the JavaScript engine that comes with HTML 5 has built-in support for encoding/decoding JSON objects. For every part placed on the design grid in the webbrowser, the parttype (ball, or block), colour (red, blue, green, yellow) and position (coordinates on the design-

grid) is entered in this JSON information. It is also possible to choose whether or not to use adhesive. By clicking the submit button, the JSON information is transferred to the webserver. Every action described in this information is related to and translated into a production step. In figure 9 the internal structure of a production step information block is given. A unique ID is followed by a capability. This is the step action required and will be tied to an equiplet capable to perform this step. The parameters give extra information about the object the action has to work on. For example in a pick and place action, the parameters will specify the coordinates of the final positions and the object that has to move to that position.

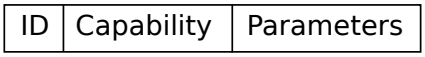


Fig. 9. Components of a step object

4.3 Webserver and Tomcat-driven Java application

The web page presented to the client is presented by a Tomcat web server. Tomcat is designed to support Java Servlets. This means that Tomcat is capable to start a Java program at the server the moment the client sends a request for a product. This Java program is capable of spawning a product agent in the Jade environment. To do this a Gateway is used in the Jade environment to achieve this functionality. This newly spawned agent will also receive the JSON information about the product to be made. From this information, the needed product steps are generated by the product agent. An overview of the connection sockets is shown in figure 10. Every product agent is capable to receive information from the Tomcat server using the Gateway Server. Every product agent can also directly send information to the Tomcat Server. This will create the possibility to inform the end-user in realtime about the progress of the production.

Product agent The product agent is created and its goal is to produce the product. Therefore it has to fulfil its sub-goals. The first sub-goal is planning the production path. This means: selecting the equiplets involved, inquire if the steps are feasible and finally scheduling the production. The next sub-goal is to guide the product along the production path and to inform the equiplet about the step or steps to perform. For every step, data aquisition of the production data is possible and should be carried out by the product agent. It depends on the equiplet agent what information will be made available.

Blackboard and timing The blackboard system as described in the architecture was implemented as actually three separate blackboards (see figure 3). This

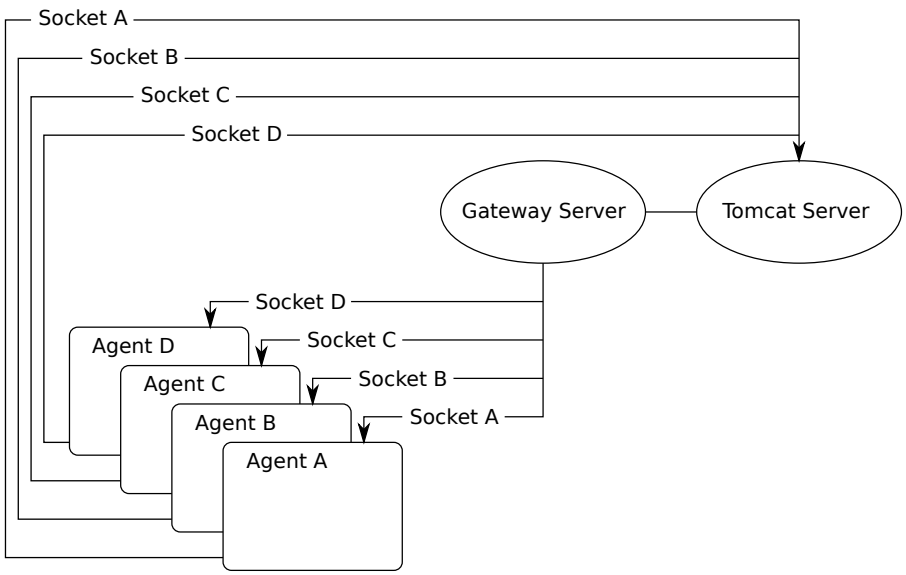


Fig. 10. Socket connections between product agents and the user interface

has to do with the fact that the performance of the system could be better and also the read and write access permissions become more clear. The BB-steps blackboard is used by the equiplet agents to announce its production steps. This information is under normal circumstances read-only for the product agents. The BB-planning blackboard is read and written by the product agents and a timing process. The information on this blackboard is the planning of timeslots or time steps for every equiplet, and a load of every equiplet.

To synchronise all agents, a timeserver has been added to the system. The scheduling is done by the product agents. Every newly arrived product agent tries to schedule itself in a way that it will not exceed its deadline. If it fails, it will ask other product agents with a later deadline to temporally give up their scheduling. Next it will try to generate new schedules for all involved agents. If successful, the new schedule will be adopted. If the scheduling fails the old schedules are restored and the new agent reports a scheduling failure.

The third blackboard in figure 3 (BB-logfile) is used to build a knowledge base about the performance of the individual equiplets and is shared among the product agents. Successful and unsuccessful steps are reported in this blackboard by products agents. This blackboard serves as an extra check when the product agent is planning the set of equiplets to be used for a certain product. The higher the failure rate of a certain equiplet, the more it will be avoided by the product agents. This failure rate can be reset after repair or adjustment of an equiplet.

Equiplet agent The equiplet agent is also implemented as a Jade agent and it is the interface to the underlying software and hardware. It depends on the

front-end of the equiplet what modules are available. The equiplet agent is also the interface to the product agent. Both types of agents live in Jade containers and can communicate with each other. The communication between the product agents and the equiplet agents as well as other product agents is FIPA-based. FIPA is an acronym for Foundation for Intelligent Physical Agents and the foundation developed a standard for inter-agent communication. The Jade platform is FIPA-compliant. For the implementation of the blackboard, Open BBS has been chosen. This Java-based blackboard was easy to integrate in the Jade environment; it was open-source and tests proved that it performed well enough for our grid.

The equiplet agent will translate the production steps in front-end-specific sub-steps. A pick-and-place action is composed of movements and control of a vacuum pincer to pick the objects involved. The movements and commands are sent to the ROS-layer that will control the hardware and the commands are actually carried out by the connected hardware.

5 Results

The research done so far for this agent-based production system had several milestones. The first milestone was the proof of concept given by a simulation of the multiagent system as described in [5]. In that system the product agents planned their production path along equiplet agents that used timing delays to mimic the production steps. The equiplet agent was not combined with the equiplet hardware. The next milestone was the implementation of a reliable and fast scheduling algorithm as described in [6]. The third step was integrating the MAS with the ROS-based equiplet in the system, so the integration with real equiplet hardware has been accomplished [12]. The latest step is described in this paper. A web front-end has been built to specify the product to be produced. At this moment the given 2D examples can be executed on the three available equiplets. So the total chain from design to production is working. In figure 11 a design in the paint application of WIMP is made. In figure 12 the result of this product is shown. Though this example still is very simple, it shows that the multiagent system is working to our expectations. The 3D example is already implemented at the MAS level and ROS level. The equiplet front-end to perform these steps is under development as a glue dispenser and an extra degree of freedom (rotation capability around the z-axis) of the pick and place robot is needed. However using a dummy equiplet (as in the earlier developed simulation) shows that the software is working to our expectations. This also includes an error recovery system.

6 Related work

The concept of using agents for production is not new. Among others a multiagent-based production system has also been developed by Jennings and Bussmann [3][4]. Jennings and Bussmann introduce the concept of a product agent, in their

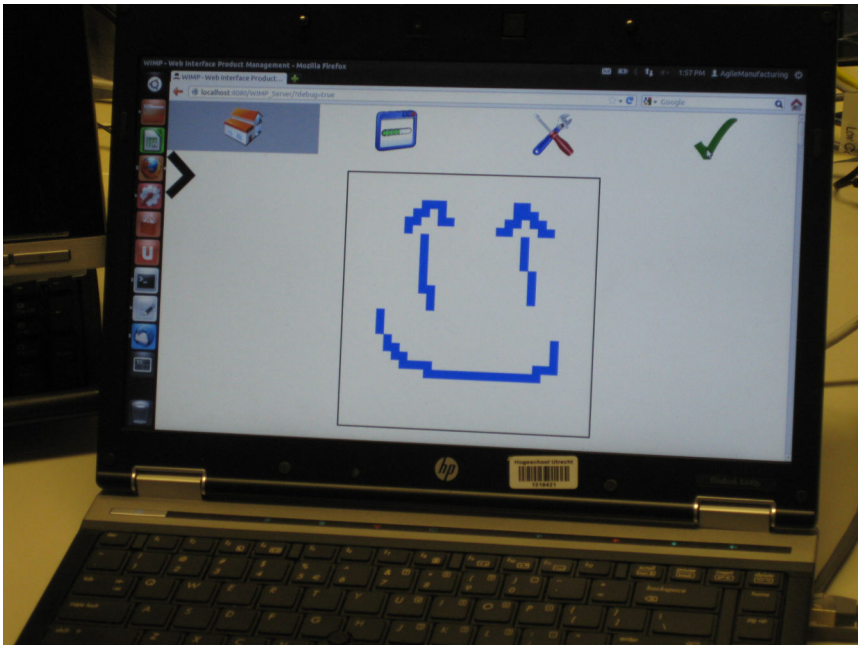


Fig. 11. WIMP paint design

terms workpiece agents, during the production. Their system focuses on reliability and minimizing downtime in a production line. This approach is used in the production of cylinder heads in car manufacturing. The roles of the agents in this production system differ from our approach. This has to do with the fact that Jennings and Bussmann use agent technology in a standard pipeline-based production system and the main purpose was to minimise the downtime of this production system. Their agents do not perform individual product logging and only play a role in the production phase. In our approach the product logging is done by the product agent for every single product and could be the basis of the other roles of the product agent in other parts of the life cycle. In the model presented by Jennings and Bussmann the workpiece agent is not so much involved in production details as the product agent in our model. Another big difference is also that our model is end-user driven.

In the field of agent-based production there are several other important publications. Paolucci and Sacile[8] give an extensive overview of what has been done. Their work focuses on simulation as well as production scheduling and control. The main purpose to use agents in [8] is agile production and making complex production tasks possible by using a multi-agent system. Agents are also introduced to deliver a flexible and scalable alternative for MES for small production companies. The roles of the agents in their overview are quite diverse. In simulations agents play the role of active entities in the production. In production scheduling and control agents support or replace human operators.

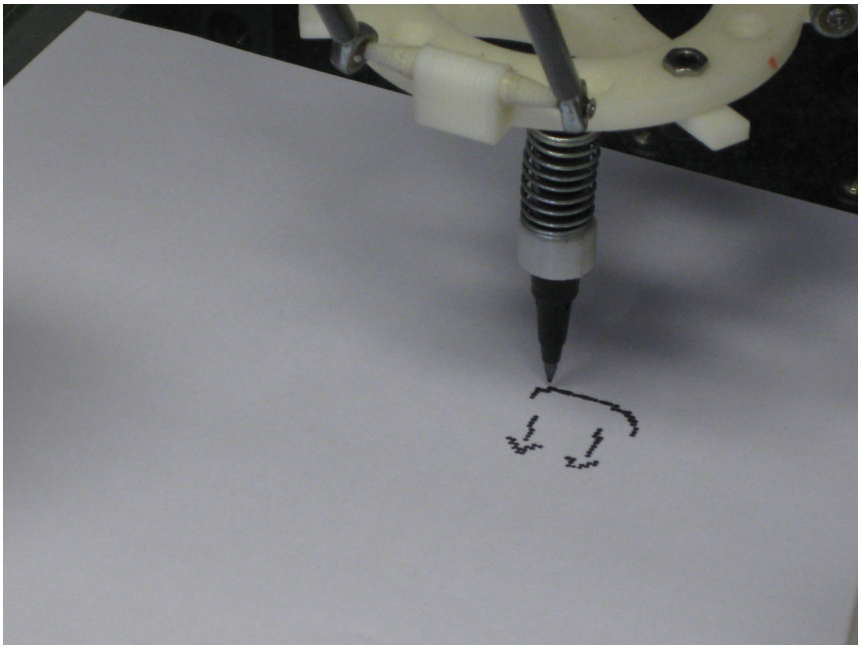


Fig. 12. Resulting product on the equiplet

Agent technology is used in parts or subsystems of the manufacturing process. We on the contrary based the manufacturing process as a whole on agent technology and we have developed a production paradigm based on agent technology in combination with a manufacturing grid. This model uses only two types of agents and focuses on agile multiparallel production. The design and implementation of the production platforms and the idea to build a manufacturing grid can be found in Puik[9]. After production the product agents can be embedded, if possible, in the product itself. In [7] the role of product agents in the whole life cycle of a product is discussed.

The term industrial internet [2] is used to describe the possibilities of interconnected machinery, sensors and devices that can be used to enhance production and solving emergent problem on the fly. Research in this field is related to our research. The approach we used however is purely based on the aforementioned cheap reconfigurable equiplets. The introduction of agent technology opens possibilities that go beyond the production phase, as the product agent can play an important role in other parts of the life cycle of a product.

7 Discussion and future work

The production approach described here is also applicable to a hybrid system containing human actors as parts of the production system. In this situation

human workers take the position of the equiplets. The production steps for a certain product should be translated to human-readable instructions and humans perform the actual production steps. In that model the equiplet agent carries out this translation so the MAS layer is still intact. This approach is useful in the situation where the production tasks are too complicated for an equiplet to be performed, but it can also help in the situation where a new equiplet front-end has to be developed.

Standard mass production always has the risk of overproduction, especially when new products arrive from other sources offering better performance or a lower price. In the concept of lean manufacturing, this kind of waste should be avoided by so-called pull-driven production. This means that a product will only be made if an end-user is asking for it. This is exactly what has been accomplished in the manufacturing system described in this paper.

For transport of the products between equiplets, automated guided vehicles (AGV) are being developed. However, the use of AGVs is not implemented yet, but the transport between equiplets can be seen as a step needed in the sequence of steps to make a product. This means that from the point of view of the product agent, an AGV is just another equiplet, offering the product transport step fitting in the total sequence of steps needed for manufacturing the product. There is a difference however. The AGV is reserved for a whole sequence of steps, while equiplets are reserved for just a single step or a set of steps if these steps are consecutive and can be realised by the same equiplet.

8 Conclusion

In this paper we described a real production system that has been built as a proof of concept. All software used is based on open standards. Further research on the manufacturing of products with a higher complexity must be done, however the basic techniques for the implementation proved to work.

The grid is capable to produce several different products in parallel and every product has its own unique production log generated by and embedded in the product agent. This product agent can play an important role in the other parts of the life-cycle of the product. When a product will be disassembled the product agent carries important information about the sub-parts of the product. This can be useful for recycling and reuse of sub-parts.

References

1. Bordini, N., Dastani, M., Dix, J., Seghrouchni, A.E.F.: Multi-Agent Programming. Springer (2005)
2. Bruner, J.: http://radar.oreilly.com/2013/01/defining-the-industrial-internet.html (2013)
3. Bussmann, S., Jennings, N., Wooldridge, M.: Multiagent Systems for Manufacturing Control. Springer-Verlag, Berlin Heidelberg (2004)
4. Jennings, N., Bussmann, S.: Agent-based control system. IEEE Control Systems Magazine (Vol 23 nr.3), 61–74 (2003)

5. Moergestel, L.v., Meyer, J.-J., Puik, E., Telgen, D.: Decentralized autonomous-agent-based infrastructure for agile multiparallel manufacturing. ISADS 2011 proceedings pp. 281–288 (2011)
6. Moergestel, L.v., Meyer, J.-J., Puik, E., Telgen, D.: Production scheduling in an agile agent-based production grid. IAT 2012 proceedings pp. 293–298 (2012)
7. Moergestel, L.v., Meyer, J.-J., Puik, E., Telgen, D.: Embedded autonomous agents in products supporting repair and recycling. Proceedings of the International Symposium on Autonomous Distributed Systems (ISADS 2013) Mexico City pp. 67–74 (2013)
8. Paolucci, M., Sacile, R.: Agent-based manufacturing and control systems : new agile manufacturing solutions for achieving peak performance. CRC Press, Boca Raton, Fla. (2005)
9. Puik, E., Moergestel, L.v.: Agile multi-parallel micro manufacturing using a grid of equiplets. IPAS 2010 proceedings pp. 271–282 (2010)
10. Quigley, M., Gerkey, B., Conley, K., Faust, J., Foote, T., Leibs, J., Berger, E., Eheeler, R., A., N.: Ros: an open source robot operating system. Open-Source Software workshop of the International Conference on Robotics and Automation (ICRA) (2009)
11. Shingo, S.: A Study of the Toyota Production System. Productivity Press (1989)
12. Telgen, D., Moergestel, L.v., Puik, E., Meyer, J.: Requirements and matching software technologies for sustainable and agile manufacturing systems. INTELLI 2013 proceedings pp. 30–35 (2013)
13. Wooldridge, M., Jennings, N.: Intelligent agents: Theory and practice. The Knowledge Engineering Review (10(2)), 115–152 (1995)

ICT and e-business development by the Ukrainian enterprises: the empirical research

Nataliia Medzhybovska

Odessa National Economic University, Odessa, Ukraine

nmedzh@oneu.edu.ua

Abstract. This paper presents the results of the research about the level of information and communication technology (ICT) implementation by the Ukrainian enterprises. We studied different Web-sites and made more detailed research about ICT implementation at the Odessa industrial enterprises. The conclusion is made that the state of Odessa e-commerce market does not correspond to the current state of ICT development, nor to the needs of the information society development in our country. The research of industrial enterprises shows insufficient use of the advantages, which can brings the effective use of ICT. Most businesses, despite the relatively high level of technical equipment, automate only a part of routine operations. Most administrative functions are performed by traditional methods, using only e-mail. Thus, the Ukrainian enterprises are facing an urgent task of the most effective use of available human and ICT potential to improve their performance and competitive position at the market.

Keywords: Industrial enterprises, business partners, information and communication technology, e-business, e-commerce, Web-site, intercompany interaction, electronic information interchange.

Key Terms. Development, industry, research.

1 Introduction

Nowadays e-business in Ukraine relates mostly to online retail shopping and to searching of relevant information and slightly extended to intra- and intercompany interaction. Moreover, the benefits of e-business are not used by all business sectors in Ukraine. The most developed in terms of Internet penetration are banks, large retail chains, high-tech enterprises, enterprises in leisure activities and entertainment, etc. Unfortunately, the industry – the main and leading sector of material production in Ukraine – lags in the context of the use of e-business tools to improve the efficiency and competitiveness of domestic enterprises.

The paper is organized as follows. Section 2 elaborates on theoretical underpinnings, and presents an overview of the related works. Section 3 provides the research methodologies and the empirical results. Section 4 concludes.

2 Related works

International researches show that "there is a growing amount of evidence from developed and developing countries that the adoption of ICTs by enterprises helps accelerate productivity grows, which is essential for supporting income and employment generation. More widespread adoption of ICTs in the productive sectors of developing countries should also accelerate innovation and thus enhance the competitive position of developing countries" [1].

A considerable amount of researches identify the benefits that brings the use of e-business tools for industrial enterprises. Researchers stated that "the potential of B2B e-commerce is not captured by merely automating document printing and mailing operations of transactions, but by encompassing all trading steps and collaboration between business partners. Firms need to realize the importance of cross-firm process integration and make determined efforts to integrate B2B e-commerce in critical business processes" [2]. Several studies concern to barriers to the implementation of B2B e-business solutions [3], the conditions necessary for their successful adaptation [4] and others.

Importantly, many researches confirm the fact that e-business in B2B is less developed compared with e-business in B2C. Scientists comprehend the reasons for this situation in the differences between B2B and B2C markets. For example, Harrison et al. identified 10 reasons, including more complex decision-making unit and associated with it increased rationality of buyers, the complexity of production, the limited number of buying units, far fewer behavioral or needs-based segments, importance of personal relationships, long-term purchases or at least purchases which are expected to be repeated over a long period of time, etc. [5]. Wright has identified the following features of B2B: decision-making structure is complex and the process involves a lot of people; decision-making could be delayed, depending on the purchase value; rational reasons for ordering; high value of product/service, contacts, projects and consulting; the final consumer probably will not be a decision-maker; since the process time increases, suppliers have the access to decision-makers [6].

In our opinion, the sufficient reason for backlog of e-business development by the Ukrainian industrial enterprises is also the backwardness of the employees who do not tend to improve their skills in the ICT field, fear of change and therefore limitation themselves to existing business practice. On the other hand, the advance of B2C sector in this area can be explained by the voice of customers who do not satisfied by the old methods of obtaining information and traditional relationships. In other words, the B2C sector is forced to respond to the market needs for maintaining its competitiveness. This need for B2B sector in Ukraine as well as for internal and intercompany automation, in our view, has not formed yet.

3 The Research

3.1 The methodologies used for e-business development research

This section presents the methodologies for the evaluation of e-business and ICT in use maturity in Ukraine on the example of the Odessa enterprises. The study was conducted in two directions.

Within the first direction, we examined the large number of enterprises in Odessa region. These companies were divided into the two groups: the industrial enterprises and companies in other business fields. We analyzed all industrial enterprises listed at the Odessa official site http://www.odessa.ua (in total 161 companies). Enterprises of other business areas were chosen randomly using the Internet handbook "World Gold Page. All Odessa" at www.mercury.odessa.ua and partly on the Odessa official site (total 664 companies).

Companies were chosen from different fields of business, they have different forms of ownership, scope of activities, etc. Choice of the specific number of companies in each area were dependent on its scale and popularity in Odessa and Odessa region. Odessa is a major business center in Ukraine, therefore the study of the level of ICT in use maturity at the Odessa enterprises is quite representative for Ukraine.

The purpose of this study is the investigation of these companies' presence in the Internet and a comparison between industrial enterprises and companies of other spheres of activity. We understand that having a Web-site does not fully indicate the involvement in e-commerce and/or e-business, but it is minimal and necessary condition for this purpose. In this context, we determine not only the existence of Web-site, but also make a detailed study of different kinds of Web-pages for named companies.

The second direction is represented by a detailed study of the level of ICT use and e-business development at the 12 industrial enterprises in Odessa, including 5 engineering companies, 3 food processors, 2 enterprises of fabricated metal products, 1 cable plant and 1 plant for the plastic products production. Research was made on the basis of employees' survey data and state statistical observation (form № 1-ICT "Information and communication technologies and e-commerce in enterprises").

On the basis of these documents, 18 indicators have been allocated, characterizing the level of ICT in use maturity and e-business development at the enterprise.

On the basis of the expert survey were identified the levels of: achievement of the objectives of information systems; the Internet use; duties automation; achievement of automation benefits. These indicators were determined based on the frequency of positive respondents' answers to the questionnaire.

Part of transactions with suppliers / customers / other organizations that are implemented via electronic information interchange (EII), as well as the part of functions that are provided by Web-site, were determined on the basis of form № 1-ICT. In this case we also used the frequency of positive answers to the relevant questions of the state statistical observation.

It should be noted that we use the term of "electronic information interchange", which assumes the use of computers and communication tools to transmit information. It includes exchange of information through the enterprise Web-site /

Web-portals including publication of information, upload / download documents, e-mails; automated data interchange systems, which exchange data in real time over a coherent structure, format and data transmission standards with minimal or no human intervention (XML, EDIFACT, etc.).

The author believes that this term adequately reflects the whole spectrum of transactions, which is mentioned in the relevant paragraphs of 1-ICT form.

Answers to questions of form № 1-ICT allow also to define the part of products which are sold via computer networks; part of material resources purchased via computer networks; availability of personal computers for administrative staff and employees; part of personal computers connected to the Internet. These parameters were defined as the quotient between the corresponding data.

The quality of the Internet connection is also determined on the basis of form № 1-ICT. The answers to this question are qualitative, so it is necessary to transform them to the quantitative form. For this reason we conducted the survey of the ICT professionals regarding the weights for different variants of Internet connection. It was found that the most relevant criteria for assessing the level of Internet connection is the connection speed, because of other communication quality parameters such as the level of support, the stability of the signal response to faults, etc. depend on the quality of a particular ISP, not on the way the Internet connection. For further calculations we used the arithmetic mean value of the respondents' answers. For each company it was compared with the maximum possible/progressive method (combination of methods) of Internet connection.

Availability of LAN, wireless LAN access, intranet and extranet were evaluated on the basis of answers to the relevant questions of form № 1-ICT. Answer “Yes” is set to 1, answer “No” is set to 0.

Assessment of the ICT in use maturity e-business development at the industrial enterprises on the above parameters was carried out using a software package for statistical analysis “Statistical Package for the Social Sciences” SPSS.

3.2 Comparative analysis of e-business development

The results of comparison of e-business adoption between Odessa industrial enterprises and enterprises of other business fields are following.

The research shows that from 161 industrial enterprises only 64 have the Web-site (40%), that indicate an insufficient level of e-business implementation in Odessa industry. We are sure that high quality Web-site can become a gateway for industrial enterprises to attract new business partners, conduct competitive procurement, provide the comprehensive information about its products, offer secure and controlled communication with business partners, etc. From the other hand, 329 companies from the 2nd group (50%) have its own Web-site.

In both cases, Web-sites often provide information in Russian language (for industrial enterprises – 89% that have Web-site, for companies in other fields – 93%). Publication information in Ukrainian is at 23% and 20% of Web-sites respectively, in English – 41% for both groups. Obviously, multilingual information is essential for companies interested in attracting a great number of customers and business partners.

We discovered the following methods of communication with companies employees through Web-site. For industry each has phone number, 72% contain the mailing address, 87,5% – email address (20 of them hosted on free hosting, that indicate a lack of attention to the positive reputation in the e-commerce market), chat, feedback service and sms are not popular (8%, 27% and 8% respectively). For companies in other business fields phone number presented at 94% of Web-sites, e-mail address – at 78%, mailing address – at 53%, feedback service – at 23%, chat – at 8%, sms – at 4%. It is logical that availability of free channels of communication for B2B sector is not as critical as for B2C e-commerce, but the presence of multiple communication channels is helpful. Furthermore, in this case it is important that communication channels are personalized (e.g., the purchasing issues should be addressed directly to the sales department, proposals from suppliers – to the procurement department, etc.). Moreover, such requests should be fixed in order to monitor the timeliness and completeness of its accomplishment.

The following results describe the quality of information presented at the companies' Web-sites. Complete information about the products is available at 90% of industrial Web-sites, partial – at 8%, no product information – at 2% of Web-sites. Complete information about the products is available at 76% of Web-sites of other business fields companies, partial – at 23%, no information available – at 1%. These figures clearly show lack of attention of Ukrainian enterprises to the use of Internet as a cheap and convenient channel for information dissemination. We believe that information about the company's products should be presented at the Web-site with the required level of detail, in some cases – with the drawings and specifications. The companies Web-site is also should have the full information about the enterprise and provide standard contracts forms, business rules, etc.

Further, the research found that industrial enterprises do not use social networking. We believe that social networks should be used by companies to maintain personal contact with business partners and to create the specific professional communities. The data regarding the presence of other business fields companies in social networks are the following: the Facebook pages have 8% of firms, accounts in Twitter – 5%, in YouTube – 4%. Moreover, for B2C companies and consumer goods' producers the social networking is critical for increasing the loyalty of existing and attracting new customers.

We also investigated the availability of information about the companies which do not have Web-sites, on the popular Ukrainian Internet portals such as Prom.ua, Businessua.com, Ua.all.biz. We proceeded from the assumption that if the company does not have a Web-site, it should have an account at the niche portals to implement at least the minimum presence in the Web.

The research shows the following data: at the Prom.ua we found 9% of the industrial enterprises, which do not have own Web-site, at the Businesua.com – 3%, at the Ua.all.biz – 26%. There are no information at these portals for 64% of industrial enterprises which do not have its own Web-site.

For companies of other business areas the data is following: 3% of companies which do not have Web-site, we found at Prom.ua, 0,6% – at Businessua.com, 6% – at Ua.all.biz. The information about 90% of companies which do not have their own Web-site, we didn't find at these portals.

Thus, the example of Odessa enterprises shows the unsatisfactory level of e-business development for both groups of enterprises. Moreover, it is the obvious gap in the e-business development by the industrial enterprises as compared to businesses of other fields. Many enterprises do not have their own Web-site, most of these companies do not even presented at the most famous Ukrainian portals. Companies that have Web-sites show its insufficient quality. Furthermore, the industry did not use the resources of social networking for disseminating information and supporting its business partners loyalty.

3.3 Research of the ICT development at the industrial enterprises

Evaluation of the ICT in use maturity at the Odessa industrial enterprises of the chosen sample allowed to make following important conclusions.

On average, the surveyed enterprises use information systems to achieve 2,4±0,2 purposes from 6, and the most popular purpose to use information systems is improving the access to information (78% of enterprises).

Internet in the surveyed companies is used to perform an average of 4,1±0,4 duties from 11. The most popular direction of Internet usage are message and document transfer to business partners (79% of enterprises), to employees and superiors (67%) and search for suppliers (62%).

On average, the surveyed enterprises automate 1,9±0,3 duties from 8. The leader among the responses is reporting (45% of enterprises).

Automation of duties allow to realize an average 4,0±0,4 purposes from 13 at the surveyed enterprises. The most commonly implemented objectives are speeding the paper documents preparing (84% of enterprises) and its transfer to employees and superiors (67%), reducing its number (79%).

Electronic information interchange with suppliers implemented by 5 enterprises from 12. These enterprises use EII on average 6,0±0,2 transactions with suppliers from 7, and all enterprises use EII for the transferring orders to suppliers, receiving electronic invoices and product information from suppliers.

Electronic information interchange with customers is implemented on the same 5 companies from 12. They realize an average of 5,9±0,2 transactions with customers from 7, and all these companies send electronic invoices and information about its products to customers.

Electronic information interchange with other organizations is executed by all enterprises. On average, they realize via EII 5,5±0,2 operations with other organizations from 7, with the most popular such as: obtaining banking and financial services (94% of enterprises), information from government agencies (90%), documents from government agencies (85%), returning of completed forms to government institutions (81%), sending or receiving data to/from government institutions (79%), sending payment orders to financial institutions (60%).

Level of EII with other organizations for all enterprises significantly exceeds the level of EII with suppliers and customers. From our point of view, this situation is caused by the availability of the relevant proposals and realized possibilities from the government institutions and banks/financial institutions (receiving and returning of

electronic documents, executing the administrative procedures, submitting proposals, using the Internet banking, etc.).

Web-site is available for 6 companies from 12, and it provides on average 2,3±0,4 functions from 6. All Web-sites contain product catalog or price list, but only one company realizes all features listed in the form № 1-ICT.

Although five companies in the study group show the automation of certain procurement and marketing functions, this automation is very limited and is implemented mainly by sending e-mail notifications to some business partners.

An indirect proof of this fact is the almost complete lack of material resources and finished products, which are purchased/sold via computer networks. Only one company 75% of its products realize through computer networks. Other company only a very small amount of its production sold this way (0.001081%), and a small part of material resources purchased via computer networks (0.01328%). Therefore, we have no reasons to report about mass and systematic electronic sales and procurement for the study group of enterprises.

An interesting analysis of the ICT equipment level for the study group of enterprises. Thus, the availability of personal computers for administrative staff indicates the gap in the level of company computerization. One company has this value higher than 5, three of them – near 1, but two – less than 0,1, others – near 0,5. The study also shows the gap between enterprises in PCs availability for employees. These values vary between 0,022 and 0,572.

It should be noted that industrial enterprises have different level of Internet connection (four of them – 100%, five – near 70%, three – only near 30%). Five enterprises apply the most advanced methods of Internet connection, including mobile communications, which increases the efficiency and timeliness of information transfer and processing, but four companies use broadband Internet connection with maximum speed 24 Mbit/s and three still use outdated analog modem or an ISDN connection.

Almost all companies have local computer network, 5 of them use a wireless LAN access. Two companies have internal computer network (intranet) and extend it for the business partners (extranet). In other words, they have the technical ability for intra- and intercompany collaboration.

Thus, the results of this research do not allow us to state the mass implementation of ICT and e-business by the Odessa industrial enterprises. In most businesses, despite the relatively high level of technical equipment, only a part of routine operations are automated. Most administrative functions are performed by traditional methods, using only e-mail.

4 Conclusions

During the conducted research, we attempted to study, firstly, the level of e-business development by the Odessa enterprises of industry and other spheres of activity, and secondly, the level of ICT usage at the Odessa industrial enterprises.

Based on this research we can make several important conclusions:

1. The state of Odessa e-business market does not correspond to the current level of ICT development, nor the needs of the building the information society in our

country. More than half of the companies do not have its own Web-sites, they are poorly represented in social networks and in niche portals on relevant topics. Detailed research of existing Web-sites showed its low quality. Web-site content and quality of its services, including on-line ordering and payment, secure communication, etc. require radical restructuring.

Comparative analysis of presence in Internet of industrial enterprises and companies of other fields of business showed the significant lag of industrial enterprises in terms of providing the information and interaction with customers and business partners. Thus, an urgent task for the Ukrainian enterprises is the most effective implementing of e-business tools into a business practice.

2. Research of the ICT equipment of the industrial enterprises showed its insufficient use, and failure to obtain the benefits that bears its effective implementation. Automation covers mostly routine paper operations, they have a very primitive mode of electronic communication with business partners and employees (often only e-mails), e-procurement and e-sales are not implemented at all, although the level of technical equipment and Internet connection are relatively high.

3. The most unfavorable situation is in the field of intercompany electronic interaction of the enterprise with its customers and business partners. The study shows almost complete absence of electronic procurement and sales. Some enterprises realize only electronic information interchange, but only with those organizations that have provide relevant technical and other possibilities for this purpose (government organizations, banks/financial institutions).

Summarizing, we can note unsatisfactory use of the ICT benefits and e-business development by the Odessa enterprises.

It should be noted that this study is only the first attempt to analyze the situation in the field of adaptation of e-business tools and ICT by the Ukrainian enterprises, so allow only to formulate the challenges in this business field. Future research should be conducted on a regular basis with the justification of representativeness (typicality) of selected companies. It also should study the companies from all regions of Ukraine, separating them by industry, type of ownership, size, etc. Also limitation of this research is the resistance of most businesses for detailed study of their activity.

References

1. The Information Economy Report. The Development Perspective. p. XX, New York and Geneva: United Nations Conference on Trade and Development (2006)
2. Claycomb, C., Iyer, K, Germain, R.L.: Predicting the level of B2B e-commerce in industrial organizations. In: Industrial Marketing Management, vol. 34, pp. 221--234 (2005)
3. Janita, I., Chong, W. K.: Barriers of B2B e-Business Adoption in Indonesian SMEs: A Literature Analysis. In: Procedia Computer Science, vol. 17, pp. 571--578 (2013)
4. Wang, S., Mao, J.-Y., Archer, N.: On the performance of B2B e-markets: An analysis of organizational capabilities and market opportunities. In: Electronic Commerce Research and Applications, vol. 11/1, pp. 59--74 (2012)
5. Harrison, M., Hague, P., Hague, N.: Why Is Business-to-Business Marketing Special? In: B2B Market Research Company. Market Research Firm. B2B International, http://www.b2binternational.com/publications/b2b-marketing (2006)
6. Wright, R.: Consumer behavior. Cengage Learning (2006)

Geospatial intelligence and data fusion techniques for sustainable development problems

Nataliia Kussul[1,2], Andrii Shelestov[1,2,4], Ruslan Basarab[1,4], Sergii Skakun[1], Olga Kussul[2] and Mykola Lavreniuk[1,3]

[1.]Space Research Institute NAS Ukraine and SSA Ukraine
(nataliia.kussul, serhiy.skakun, andrii.shelestov, basarabru)@gmail.com, nick_93@ukr.net
[2.]National Technical University of Ukraine "Kyiv Polytechnic Institute", Kyiv, Ukraine
olgakussul@gmail.com
[3.]Taras Shevchenko National University of Kyiv, Kyiv, Ukraine
[4.]National University of Life and Environmental sciences of Ukraine

Abstract. Knowledge on spatial distribution of land cover and land use is extremely important for solving applied problems in many domains such as agriculture/food security, environmental monitoring, and climate change. Geospatial data including satellite imagery play an important role since it can provide regular, consistent and objective information. Identifying geospatial patterns and quantifying changes that occur in space and time require special techniques to be exploited. These techniques are associated with the area of geospatial intelligence and deal with multi-source data fusion and exploitation of advance intelligent methods. This paper presents the use of these techniques for processing archived and up-to-date satellite imagery for large-scale land cover and crop classification in Ukraine. The main purpose of this paper is to not only show potential of geospatial intelligence, but to pay attention of educators to this extremely important area.

Keywords. Geospatial intelligence, land cover, crop mapping, image processing, satellite imagery, big data.

Key Terms. HighPerformanceComputing, MachineIntelligence, InformationTechnology, Intelligence, Data.

1 Introduction

Geospatial information is a very important source of data for distributed systems development, education, decision making and competitive business. Due to regular acquisition of satellite data all over the world for the last couple of decades as well as new communication, navigation and crowdsourcing techniques, it has become possible to monitor the current state of the large territories development, estimate trends, analyze available scenarios for future development and manage things to provide sustainability. The approach is based on modern IT, namely geospatial

intelligence [1] and data fusion [2] techniques. By geospatial intelligence we consider all aspects of geospatial data processing including intelligent methods and technologies to fuse/integrate data and products acquired by multiple heterogeneous sources using machine learning techniques and emerging big data and geo-information technologies. In this paper we exploit geospatial technique to address two important applications for Ukraine, in particular land cover/land use mapping and crop mapping. The purpose is to not only show the potential of geospatial intelligence, but to pay attention of the educators to this powerful IT and bridge the gap between market needs for such specialists and professionals.

Ukraine is one of the main crop producers in the world [3], so agricultural monitoring is a very important challenge for Ukraine. One of the most promising data sources to solve the underlined tasks at large scale is remote sensing data, namely the satellite imagery [4-12]. This is mainly due capabilities to timely acquire images and provide repeatable, continuous measurements for large territories. At present, there are only coarse-resolution satellite imagery (500 m spatial resolution), that has been utilized to derive global cropland extend, e.g. GlobCover, MODIS [13]. But, low-resolution maps always underestimate or overestimate certain land cover or crop type areas. Also several global land cover maps have been made using higher resolution data such as from Landsat-series satellites [14-15], but they are not accurate enough at regional level for Ukraine. Therefore, creation of global products, such as land cover maps and crop maps, based on high resolution satellite images (at 30 m) is very important task for sustainable economic development of Ukraine. This paper presents the results of regional retrospective high resolution land cover mapping and large scale crop mapping for Ukrainian territory using multi-temporal Landsat-4/5/7/8 images and also some supporting data and knowledge obtained during our own investigations [7-8]. The main results of the work were obtained within EC-FP7 project “Stimulating Innovation for Global Monitoring of Agriculture and its Impact on the Environment in support of GEOGLAM” (SIGMA).

2 Objective of the study and data description

The paper covers two different studies: retrospective land cover mapping and crop mapping. These two problems are solved using the same geospatial intelligence approach that encompasses the use of advanced machine learning techniques. In particular, we use a combination of unsupervised and supervised neural networks to first restore missing values in multi-temporal images, and then to provide a supervised classification with an ensemble of multilayer perceptrons (MLPs). One of the advantages of this approach is possibility for automatic processing taking into account of large amount of satellite imagery that need to be processed.

At the first study, we used atmospherically corrected Landsat-4/5/7 products to produce land cover maps for land cover change detection. This was performed for all territory of Ukraine and required processing of about 500 Landsat scenes to cover it completely for three decades: 1990s, 2000s and 2010s. Also, we manually formed training and test sets for supervised classification using the photo interpretation

method. Train and test sets were created with uniform spatial distribution over the territory of interest and proportional representation of all land cover classes, namely artificial surface, cropland, grassland, forest, bare land and water.

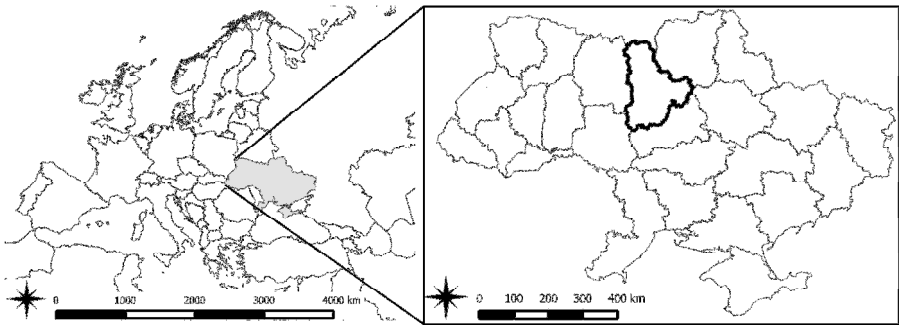


Fig. 1. Location of Ukraine and JECAM test site in Ukraine (Kyiv oblast, marked with bold boundaries.

The second study is the pilot project on large scale crop mapping for JECAM test site [16] in Ukraine for 2013 (Fig. 1). The Joint Experiment for Crop Assessment and Monitoring (JECAM) is an initiative of GEO Agriculture Monitoring Community of Practice with the intent to enhance international collaboration around agricultural monitoring towards the development of a "system of systems" to address issues associated with food security and a sustainable and profitable agricultural sector worldwide (http://www.jecam.org). The JECAM test site in Ukraine was established in 2011 and covers administrative region of Kyiv oblast with the geographic area of 28,100 km2 with almost 1.0 M ha of cropland. For large scale crop mapping over the study region we used two data sources – remote sensing images acquired by Operational Land Imager (OLI) sensor aboard Landsat-8 satellite and data acquired at ground surveys. We used Fmask algorithm for clouds detection and masking [17]. Ground surveys were conducted in June 2013 to collect the knowledge about crop types and land cover types (Fig. 2) over the interested area. In this study we used European LUCAS nomenclature as a basis for land cover / land use types.

3 Method and results

The main scientific challenges for geospatial intelligence problem solving are geospatial data fusion and correct interpretation of geospatial information. To address them for big data satellite monitoring problems we propose the novel approach, based on combination of three machine learning paradigms for geospatial information analysis: big data segmentation, neural network classification and data fusion. Data fusion is performed at the pixel and at the decision making levels. During preprocessing stage, Landsat-4/5/7 and Landsat 8 scenes were merged to multi-channel format for each path, row and date. First, we restore cloudy pixels from time-series of images using self-organizing Kohonen maps [18] and after provide

classification based on the time-series of restored images available for the certain year and required area. Classification was done by using an ensemble of neural networks (MLPs). The method of pixel and decision making level data fusion is proposed in [16].

Table 1. Accuracy comparison of Land Cover30-2010 and GlobeLand30-2010

Product	Land Cover30-2010		GlobeLand30-2010	
Class	UA, %	PA, %	UA, %	PA, %
Artificial	100	87.8	79.5	3.4
Cropland	93.5	96.2	99.4	85.3
Forest	95.4	96.2	89.9	95.9
Grassland	81.4	71.2	34.4	60.5
Bare Land	91.7	96.4	0.4	57.1
Water	99.5	99.6	96.6	99.9
Overall accuracy, %	94.7		89.7	

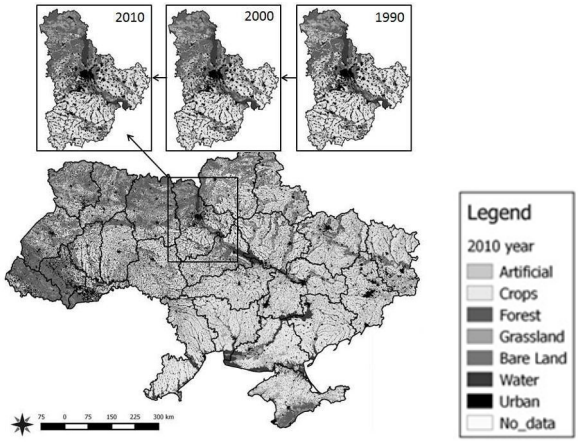


Fig. 2. The land cover map of Ukraine for 2010 year (and also land cover maps of Kyiv oblast for 2010, 2000 and 1990 years).

To estimate the accuracy of land cover classification for Ukrainian territory, we used two approaches: accuracy assessment on independent test (testing) set and comparison of the class areas in land cover with official statistics. The overall classification accuracy achieved in this study was approximately 95%. Accuracies for each individual class were more than 70%. The lowest classification accuracy was for grassland, because it is difficult to separate grassland from some of spring crops. We also compared (Table 1) our result, taken for Ukraine with global land cover map GlobeLand30-2010 at 30 m resolution. The overall classification accuracy of our land cover map was 5% higher than GlobeLand30-2010. Also accuracy of grassland from

our maps was +10% (producer accuracy, PA) and +45% (user accuracy, UA) [19] better than GlobeLand30-2010. Our final land cover map is shown at Fig. 2.

Table 2. Classification results

No	Class	PA, %	UA, %
1	Artificial	100.0	97.9
2	Winter wheat	95.7	91.8
3	Winter rapeseed	93.5	99.4
4	Spring crops	40.6	34.6
5	Maize	90.5	86.8
6	Sugar beet	94.9	89.6
7	Sunflower	84.1	85.4
8	Soybeans	69.7	77.1
9	Other cereals	70.9	78.0
10	Forest	96.9	92.9
11	Grassland	91.0	89.0
12	Bare land	86.7	99.0
13	Water	100.0	98.1

3.1 Large scale crop mapping

The use of multi-temporal Landsat-8 imagery and an ensemble of MLP classifiers allowed us to achieve overall accuracy of slightly over 85% (Table 2) which is considered as target accuracy for agriculture applications.

Target accuracy of 85% was also achieved for winter wheat, winter rapeseed, maize and sugar beet. For the spring crops, sunflower and soybeans the accuracy is less, than 85%. Soybeans is the least discriminated summer crop with main confusion with maize. In particular, almost 61% of commission error and 71% of omission error was due to confusion with maize. All non-agriculture classes including forest and grassland yielded PA and UA of more than 85%. The final classification map is shown in Fig. 3.

Comparison of official statistics and crop area estimates derived from Landsat-8 imagery for Kyiv region described at the Table 3.

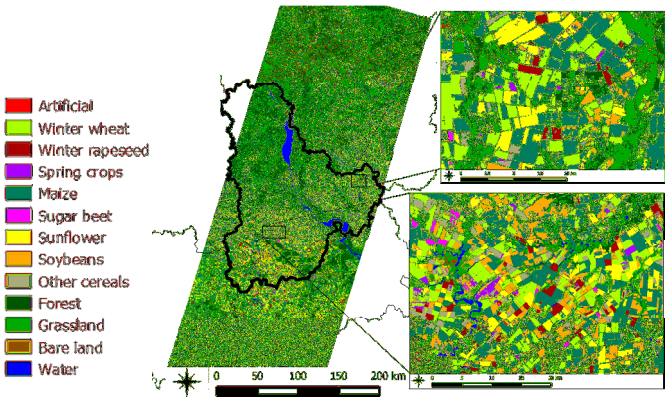


Fig. 3. Final crop map obtained by classifying multi-temporal Landsat-8 imagery.

Table 3. Comparison of official statistics and crop areas derived from Landsat-8 imagery

Class no.	Class	Crop area: official statistics, x 1000, ha	Crop area: Landsat-8 derived, x 1000, ha	Relative error, %
2	Winter wheat	187.3	184.5	-1.5
3	Winter rapeseed	46.7	59.9	28.3
5	Maize	291.7	342.4	17.4
6	Sugar beet	15.5	11.2	-27.9
7	Sunflower	108.2	117.6	8.7
8	Soybeans	145.9	168.5	15.5

4 Application in education process

As well as geospatial intelligence is one of the emerging areas of data science, we actively use it in education process. Developed approach to land cover and crop mapping is actively used for education purposes. We incorporate these topics (geospatial intelligence methods and developed software) into a master and PhD program of “Ecological and economic monitoring” specialization at the National University of Life and Environmental Sciences of Ukraine with the main focus on big geospatial data processing and satellite data analysis.

Also we are actively trying to implement project based education, involving students into scientific projects . Some methods of data fusion are included into laboratory works on intelligent computations. Master and PhD student fulfill their qualification diplomas within international projects. According to our experience more attention should be paid on geospatial data processing and intelligent

computations within Bachelor programs on Computer Science in Life Science universities.

5 Conclusions

This paper presents a novel approach for satellite monitoring based on big geospatial data analysis. The main idea of the proposed geospatial intelligence approach is the use of supervised neural networks in order to classify multi-temporal optical satellite images with the presence of missing data. A supervised classification was performed with the use of ensemble of MLP classifiers to create such global products as retrospective land cover and crop maps for the whole territory of Ukraine. Proposed approach allowed us to achieve the overall classification accuracy of 95% for three different time periods (1990, 2000 and 2010) and improve quality of maps comparing to other land cover maps available for Ukraine at 30 m spatial resolution, namely GlobeLand30-2010. The same approach was successfully applied for the JECAM test site in Ukraine for large area crop mapping.

Now geospatial intelligence is a hot topic in big data analysis, but we observe the lack of experts in the area. Therefore, we would like to pay attention of the IT educators to the gap and build a roadmap to fill it.

References

1. Bacastow, T.S., Bellafiore, D.J.: Redefining geospatial intelligence.American Intelligence Journal, pp. 38-40. (2009)
2. Hall, D., Llinas, J.: Multisensor Data Fusion. CRC Press. 568 p. ISBN 9781420038545. (June 20, 2001)
3. Crop monitor. February 2015 Maps and Charts, http://geoglam-crop-monitor.org/pages/monthlyreport.php?id=201502&type=WT (April 19, 2015)
4. Shelestov, A.Yu., Kravchenko, A.N., Skakun, S.V., Voloshin, S.V., Kussul, N.N.: Geospatial information system for agricultural monitoring. Cybernetics and Systems Analysis, vol. 49, no. 1, pp. 124-132. (2013)
5. Kussul, N., Shelestov, A., Skakun, S., Li, G., Kussul, O., Xie, J.: Service-oriented infrastructure for flood mapping using optical and SAR satellite data. International Journal of Digital Earth, vol. 7, no. 10, pp. 829 – 845. (2014)
6. Kogan, F., Kussul, N., Adamenko, T., Skakun, S., Kravchenko, O., Kryvobok, O., Shelestov, A., Kolotii, A., Kussul, O., Lavrenyuk, A.: Winter wheat yield forecasting: A comparative analysis of results of regression and biophysical models. Journal of Automation and Information Sciences, vol. 45, no. 6, pp. 68-81. (2013)
7. Gallego, J., Kussul, N., Skakun, S., Kravchenko, O., Shelestov, A., Kussul, O.: Efficiency assessment of using satellite data for crop area estimation in Ukraine. International Journal of Applied Earth Observation and Geoinformation, no. 29, pp. 22-30. (2014)

8. Gallego, J., Kravchenko, A.N., Kussul, N.N., Shelestov, A.Yu., Grypych, Yu.A.: Efficiency assessment of different approaches to crop classification based on satellite and ground observations. Journal of Automation and Information Sciences, vol. 44, no. 5, pp. 67-80. (2012)
9. Kussul, O., Kussul, N., Skakun, S., Kravchenko, O., Shelestov, A., Kolotii, A.: Assessment of relative efficiency of using MODIS data to winter wheat yield forecasting in Ukraine. 2013 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2013), pp. 3235–3238. (2013)
10. Kussul, N., Shelestov, A., Skakun, S., Li, G., Kussul, O.: The Wide Area Grid Testbed for Flood Monitoring Using Earth Observation Data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 5, no. 6, pp. 1746-1751. (2012)
11. Skakun, S., Kussul, N., Shelestov, A., Kussul O.: Flood Hazard and Flood Risk Assessment Using a Time Series of Satellite Images: A Case Study in Namibia. Risk Analysis, vol. 34, no. 8, pp. 1521-1537. (2014)
12. Kogan, F., Kussul, N., Adamenko, T., Kussul, O., Lavrenyuk, A.: Winter wheat yield forecasting in Ukraine based on Earth observation, meteorological data and biophysical models. International Journal of Applied Earth Observation and Geoinformation, vol. 23, pp. 192-203. (2013)
13. MODIS Data Products Table. Product MCD12Q1, https://lpdaac.usgs.gov/products/modis_products_table/mcd12q1 (April 19, 2015)
14. Geoportal Openlandservice, http://www.globallandcover.com (April 19, 2015)
15. Geoportal ESA CCI Land Cover products: a new generation of satellite-derived global land cover products, http://maps.elie.ucl.ac.be/CCI/viewer/index.php (April 19, 2015)
16. Kussul, N., Skakun, S., Shelestov, A., Kussul, O.: The use of satellite SAR imagery to crop classification in Ukraine within JECAM project. IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2014). (2014)
17. Zhu, Z., Woodcock, C.E.: Object-based cloud and cloud shadow detection in Landsat imagery. Remote Sensing of Environment, vol. 118, pp. 83–94. doi:10.1016/j.rse.2011.10.028 (2012)
18. Skakun, S., Basarab, R.: Reconstruction of Missing Data in Time-Series of Optical Satellite Images Using Self-Organizing Kohonen Maps. Journal of Automation and Information Sciences, vol. 46, no. 12, pp. 19-26. (2014)
19. Story, M., Russell, G.: Congalton Accuracy assessment - A user's perspective. Photogrammetric Engineering and Remote Sensing, vol. 52, no. 3, pp. 397-399. (March 1986)

Risk Assessment of Use of the Dnieper Cascade Hydropower Plants

Andriy Skrypnyk[1], OlhaHoliachuk[1]

[1] National University of Life and Environmental Sciences of Ukraine
avskripnik@ukr.net, olia_ailo34567@ukr.net

Abstract. In this article we wish to evaluate efficiency of use of Dnieper cascade hydropower plants on the basis of common approaches to environmental management. We evaluate the efficiency of use the flooded areas of the hydropower station in agriculture. Assessment of the man-made risks includes evaluation of static (regular maintenance of dams) and stochastic (probability of artificial tsunami) components. According to the world statistics of disasters caused by dam reservoirs, the probability of man-made tsunami is estimated around 0.01%. Using this rate of probability we can state that expected losses van be 5% of the confidence level. Dnieper reservoirs ranking on the degree of energy risk (the possibility of man-made tsunami generation) was made.

Keyword. risk assessment**,** hydropower plan**,** electricity, agriculture, environmental management.

Key Terms. MathematicalModel, Data, Environment, Infrastructure, Development.

1 Introduction

Before the era of nuclear power, contribution of hydropower in the energy balance of the former Soviet Union was considered indisputable. Thus the negative effects associated with the creation of reservoirs on the plains were not taken into account e.g. flooding of large areas, destruction of towns and historic monuments, increase of the risk of man-made disasters. But time passed and in 1970s in Ukraine were built several nuclear power plants and as a result appeared the need to develop solar, wind and bioenergy and it led to decrease of the share of electricity generation by hydropower plants to 5-7%. Over the past decade, the agricultural sector of the Ukrainian economy has become one of the major players in the global food market and agricultural export of the country has become one of the landmarks of the national economic development. That is why there is an urgent need to use territory of the cascade of Dnieper reservoirsfor agricultural purpose. However, beside inappropriate use of land resources [16] and deterioration of the quality of water resources there is a high risk of man-made disasters which can be caused by the functioning of the Dnieper cascade hydropower plants.

A. Pigou [11], P.Samuelson [15], R. Coase [12] presented classical approaches to exploration of the impact of externalities on economic performance (environmental management). The main idea of this approach is that the price of products (in this case electricity) does not respond the social price paid by people for violations of the environment [13] and therefore assessment of economic growth should be calculated taking into account the price of deterioration of the environment [1, 2] .

English researchers proposed classical approach to the exploration of the causes of destruction of dams, they assert the classical definition of the threats which are connected with creation of artificial reservoirs [16]. In Great Britain all the artificial reservoirs (more than 25 000 cubic meters of the size 100m * 100m * 2.5m) were under the control of local authorities and then the responsibility to control artificial reservoirs was transferred to National Environment Agency.

For comparison, Kyiv reservoir has a volume of 3.73 billion cubic meters and it is placed above the level of many districts of Kyiv [9]. Kurenevka tragedy which happened in 1961 showed that even not significant in volume reservoirs (600 000 cubic meters - 400m * 400m * 3,75m) can be extremely dangerous if they are placed above the level of the nearby territories and can lead to generation of artificial tsunami [9]. During World War II in parts of the Dnieper River below the Dnieper dam the retreating Soviet army tried to destroy the dam. The man-made disaster led to the flood victims among whom were citizens of Zaporozhe and coastal villages and soldiers of the Soviet Army (about 100 000 people) [6].

Researchers emphasize the negative effects of the creation and functioning of the Dniper reservoirs, besides flooding of large territories the negative effects concern a change of hydrological, hydro chemical and hydro biological regimes and slowing of water circulation [3, 6, 7]. In general, there is a great number of scientific papers on significant negative effects connected with creation of the Dnieper reservoirs for the environment of Dnieper, in particular, and for the economy of Ukraine in general. But the issue of quantitative estimation of possible losses caused by artificial tsunami has got little attention among researchers.

In this article we wish to explore a comprehensive risk assessment of further functioning of Dnieper hydroelectric cascade considering alternative options of usage of flooded areas and possible losses connected with future functioning of reservoirs and to develop the methodology of losses assessment connected with destruction of dams reservoirs.

2 Characteristics of Flooded Areas and Options of Alternative Exploitation

As we already mentioned, in twenty-first century the hydroelectric power generation ceased to be a decisive factor in the energy balance. The GDP growth in 2000-2007 was not connected with an increase in production of electricity. During this period was an increase in production of cereals for which the usage of electricity was minimal. It is difficult to estimate the total social costs of flooded areas, and overall benefits from the functioning of large reservoirs of water. In addition, it is difficult to assess losses

connected with deterioration of water quality due to the lack of flow. There are a lot of other aspects that do not prove the necessity and efficiency of the functioning of reservoirs. However, we will focus on two main aspects: 1) alternative usage of reservoir areas in agricultural production; 2) level of risk connected with further exploitation of reservoirs (dams of the Dnieper reservoirs). Dnieper cascade hydroelectric station was built during the period of planned economy, the first dam was built on Dnieper in 1927 (Zaporozhe) and the last in 1976 (Kanev). General characteristics of reservoirs and power plants are presented in Table 1. The total area of Dnieper reservoirs is 6.9 thousand sq. km, 1.1% of the territory of Ukraine (Table 1). But if we take into account that the territory near rivers area was always the most fertile for agricultural sector, it is necessary to assess the share of reservoirs in the volume of agricultural land, which is 1,7%. Not all agricultural lands are fertile that is why the factual area which is used in agricultural sectoris about 27 million hectares (270 thousand sq. km) with a standard deviation 0.8 million hectares [4]. In this case, the share of the Dnieper reservoirs increases up to 2.6% of the area used for agriculture.

Table 1. Structure of Land Resources of Ukraine

The main types of land and economic activity	Total area	
	Thousand sq. km	% of total area
Agricultural land	415	68.8
Forests	106	17.6
Built-up areas	38	6.3
Territories covered by surface water (Dnieper reservoirs)	24(6.9)	4.0(1.1)
Unsuitable land for agricultural production	21	3.3
Total (territory of Ukraine)	604	100.0

Source:[4]

Compare the coast of total volume of products available through agricultural production from flooded areas after the creation reservoirs, and the cost of electricity generated by hydroelectric Dnieper cascade. General characteristics of reservoirs and their electricity generation capacity is presented in Table 2. From the total area of reservoirs we extracted the natural are of water surface using natural characteristics of the Dnieper and obtained the size of flooded areas which potentially could be used in agricultural sector.

The area of flooded territory is 6 thousand. sq. km. Dnieper cascade which consists of six hydroelectric power plant produce 10 billion kw * hr. per year, (40% are produced by Dnieper, 15% by Kremenchuk and15% by Kakhovska, 13% by Dniprodzerzhynsk, 10% by Kaniv, 7% by Kyiv. Dnieper hydropower station (HPS) has the best ratio of natural areas to the area of the reservoir - 38% and the worst ration has Kyiv HPS - 5%.

Table 2. Main Features of Reservoirs

	The average depth, m	The height of the dam, m	Volume, million cubic m	The potential energy of man-made tsunami, J $*10^{14}$	Area, square km	The natural area, square km	Flooded area, square km	Capacity, MW	Average annual production, mln kW • h
	1	2	3	4	5	6	7	8	9
Kyiv	4.0	11.5	3730	3.4	922	44	878	408.5	683
Kaniv	4.3	10.5	2500	2.04	581	110.7	470.3	444	972
Kremenchuk	6.0	17	13520	19.8	2252	166.5	2085.5	632.9	1506
Dniprodzerginsk	4.3	12.6	2460	2.5	567	102.6	464.4	352	1328
Dnipro	8.1	35.4	3320	10.2	410	154.8	255.2	1569	4008
Kakhovske	8.4	16	18180	21.04	2155	276	1879	351	1489
					6887		6032		9986

Source: [3;10]

We will explore the possibilities of obtaining agricultural production in flooded areas, we will start from evaluation of the efficiency of agricultural areas during last four years. Due to the high risk of the agricultural sector we use averaged indicator of efficiency during four last the years (Table 3).We obtained the indicator of efficiency of the usage of 1 thousand square kilometer of flooded areas which is equal o 0.89 billion UAH (prices of 2012), with a standard error of 0.03 billion. This means that we can get agricultural products at total value of 5.4 billion UAH (with a standard error of 0.2 billion) from flooded territories which are under Dnieper cascade

Table 3. General characteristics of the agricultural sector for the period 2010-2013

	2010	2011	2012	2013	$\bar{x}(\sigma(\bar{x}))$
Volume of production (billion UAH)*	194.9	233.7	223.2	252.9	226.2(10.3)
Area (sq. km)	246.4	247.1	261.3	262.0	254.2(3.7)
Agriculture return (%)	21.1	27.0	20.5	11.2	20.0(2.8)

* prices of 2012

Source: AgriculturalUkraine 2013 / Kyiv.-2014-p.187-200.

The value of electricity produced during a year and financial value of potential agricultural products are presented inTable 4.

We introduce the concept of efficiency of areas of separated reservoirs as the ratio of the value of the annual volume of electricity produced to the potential value of agricultural products that can be grown on flooded areas.

Table 4. The effeciency of the flooded areas in monetary terms

Name of reservoir	Flooded area, square km.	Output of agricultural products on the flooded areas, bln. USD.	Price of electricity produced, bln. USD. (VAT included)	The efficiency of the flooded areas, %
Kyiv	878	0.78	0.22	28.2
Kaniv	470.3	0.42	0.31	74.8
Kremenchuk	2085.5	1.86	0.49	26.1
Dniprodzerginsk	464.4	0.41	0.43	103.5
Dnipro	255.2	0.23	1.29	568.6
Kakhovske	1879	1.67	0.48	28.7
Total	6032	5.37	3.22	60

Source: own calculations

The total amount of the value of electricity produced is significantly less than the potential value of agricultural products that can be grown on the flooded areas, the value of electricity is only 60% of the potential value of grown agricultural products relative to the average indicator of Ukraine agricultural productivity. Graphical representation of efficiency for certain reservoirs is shown in Figure 1.

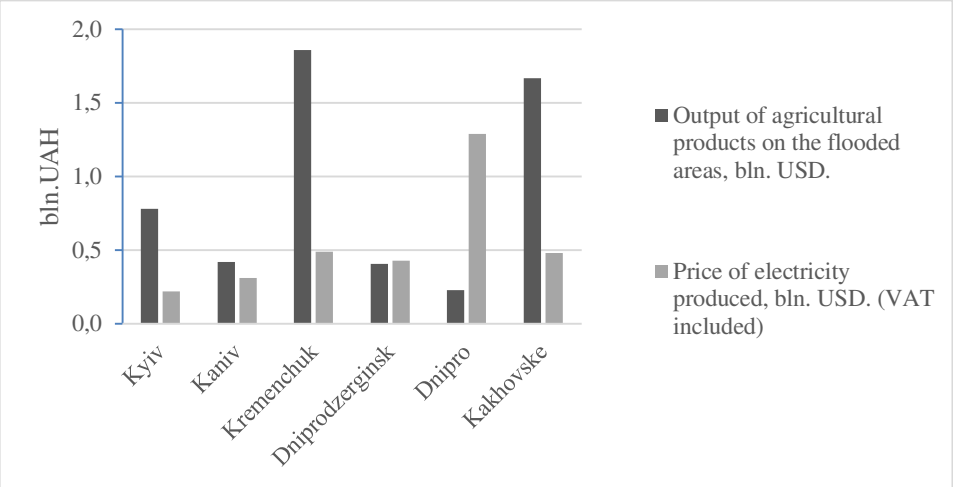


Fig.1. Comparison of possible income from agricultural production and power generation

Source: own calculations on base table 4

Data in Table 4 show that the efficiency of the flooded areas is significantly different for different reservoirs. The most effective reservoir is Dnieper HPS, because it was built in the place where the flow of Dnieper is rather fast (significant differences in levels). Further construction of power hydro stations led to the flooding of large areas that would have greater value if they were used in agricultural sector.

3 Risks Evaluation of Further Dnieper Cascade Functioning

All possible losses connected with functioning of reservoirs are not limited to the wastage of flooded areas. The general scheme of the risks evaluation of further functioning of reservoirs is presented in Figure 2. They can be divided into three groups: economic, technological and environmental.

We made an attempt to assess the expected total annual losses $\overline{L}$ which consist of economical - L_{ek} ; ecological - L_{ekol} ; and technological - L_t :

$$\overline{L} = L_{ek} + L_{ekol} + L_t \tag{1}$$

In the first approximation economic losses are equal to the difference between the price of potential agricultural products V_{ap} and the value of producing electric energy V_e:

$$L_{ek} = V_{ap} - V_e \tag{2}$$

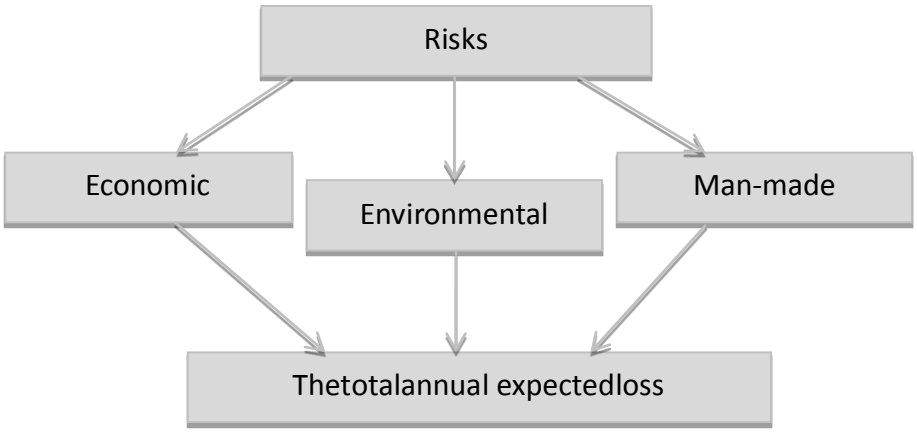


Fig. 2. Model of possible risks of functioning of Dnieper reservoirs

Environmental risk in a first approximation must be evaluated on the basis of cost of measures aimed to bring the mass of water in the reservoir (with absence of flow) to state of the river water.

The most difficult to evaluate are technological (man-made) risks, which present both static (regular repair of dams, measures aimed to support state reservoirs) and stochastic components. The latter is relevant to the possibility of artificial tsunami due to partial or complete destruction of the dam. Taking into account the global statistics the probability of the destruction of the dam is evaluated around 0.01% [14]. At first glance it is a small probability and it seems that is can be ignored, but the evaluation of the probability of depressurization of the reactor of Chernobyl type was considered lower for two orders of magnitude (0.0001%), which did not prevent this to happen. We evaluate the risk of man-made reservoir functioning for each reservoir. The potential energy that depends of the height of the dam and of the volume of reservoir after the destruction of the dam creates an artificial tsunami (Table 2). We wish to explore the least effective case (Table 4) and the most dangerous in terms of potential losses– Kyiv reservoir.

The approximate evaluation of the power of the artificial tsunami in case of destruction of the dam of Kyiv HPS can be calculated on the basis of the potential energy of water masses and sludge. The volume of the Kyiv reservoir is 3730 million ton. (Table 2) to which we add 90 million tons of radioactive sludge [8]. The average depth of reservoir is 4m and the average height of dam is 11.5m, dam reservoir center of gravity is situated at a height of 9.5 m according to the water level of the Dnieper River after the dam. That is why the potential energy of artificial tsunami that threatens Kyiv is:

$$
\begin{aligned}
&E_{ц} = m \cdot \Delta h \cdot g \approx 3730 \cdot 10^{9} \cdot 9.81 \cdot 11.5 \approx 4.2 \cdot 10^{14} J \\
&m = (3.73 + 0.09) \cdot 10^{12} kg; \\
&\Delta h = 11.5м - 2м = 9.5m; \\
&g = 9.81 m/s^{2}
\end{aligned}
\tag{3}
$$

According to the energetic characteristics the potential tsunami that threatens Kyiv is equal to five nuclear charges dropped during the Second World War on Hiroshima (15-20 kt. TNT) [15]. Of course, the shock effect of nuclear explosion and artificial tsunamis is difficult to compare because the shock wave in the first case expands at speed exceeding the speed of sound and artificial tsunami speed is determined by the depth of the Dnieper, and taking into consideration the depth of Dniper the speed will not exceed 30 km / h.).

The situation is complicated by the presence of 90 million tons of radioactive sludge at the bottom of the reservoir, the presence of which can contribute significantly to strengthening of the effects of artificial tsunami and the risk of radioactive contamination of the Dnieper and coastal areas to Kanev reservoir. In the case of this scenario, 10% of Kyiv may be contaminated [9].

Similar characteristics are calculated for each of the reservoirs (Table 2). Kremenchuk and Kakhovka reservoirs have the highest level of risk connected with emergance of artificial tsunami, it can be explained by volume of the accumulated water.

Losses caused by artificial tsunami in certain time *t* due to the violation of the integrity of the dam are proportional to the product of the tsunami energy (E_{ts}) and cost values (urban infrastructure) located in the area of artificial tsunami(S_{ots}):

$$L_t = k \cdot E_{ts} \cdot S_{ots} \tag{4}$$

where, k – coefficient of dimension J^{-1}, which can be determined only empirically.

The expected losses:

$$\bar{L}_t = p \cdot L_t \tag{5}$$

Variance:

$$\sigma^2 = p \cdot L_t^2 \cdot (1 - p) \approx p \cdot L_t^2 \Rightarrow \sigma = L_t \sqrt{p} \tag{6}$$

Losses in confidence level α - $L_\alpha (p(L \geq L_\alpha) = \alpha)$ [13]:

$$L_\alpha = \bar{L}_t + x_\alpha \cdot L_t \sqrt{p} = L_t (p + x_\alpha \sqrt{p}), \tag{7}$$

where x_α -quantile of the normal distribution.

We make an assessment of potential losses of Kyiv which can be caused by the potential of artificial tsunami concentrated in the Kiev reservoir.

Up to 10% of the houses located in Kyiv according to the evaluation of hydrologists are under the tsunami risk. The volume of living area in houses in Kyiv is 62.2 million square meters [5]. The cost of 10% of Kiev buildings, at an average price of 0.5 thousand dollars per sq. m, is 3.1 billion USD. Hence, the expected losses for a given probability of violating the integrity of the dam is $3 \cdot 10^5$ dollars. Losses in confidence level α:

$$\begin{aligned} L_\alpha &= \overline{L}_t + x_\alpha \cdot L_t \sqrt{p} = L_t (p + x_\alpha \sqrt{p}) = \\ &= 3.1 \cdot 10^9 (10^{-4} + 1.65 \cdot \sqrt{10^{-4}}) = \\ &= 3.1 \cdot 10^9 \cdot 0.0166 = 5.1 \cdot 10^7 \end{aligned} \tag{8}$$

This means that the annual potential losses from the use of the Kiev reservoir taking into account the risk of man-made tsunami are near 51 million USD.

After analyzing potential threats and possible damage, which can be caused by artificial tsunami in Kyiv we cannot propose the immediate dismantling of all the dams on the river Dnieper. The data in Table 2 on artificial potential energy of the tsunami should be supplemented by information connected with potential losses according expression (8). There must be made a forecast of losses caused by the destruction of the reservoirs. After all the calculations, we can evaluate the hazard rank of every reservoir and thus offer the procedure of their disassembling in order to restore the natural state of the Dnieper.

4 Conclusions

New information technologies and development of the theory of environmental management leads to a revision of the main concepts of the planned economy. Thus it leads to the change of our view on necessity and efficiency of functioning of hydropower stations. We analyzed the energetic efficiency of certain reservoirs on the basis of an alternative use of the flooded territory in agriculture. Energy efficiency of different reservoirs is rather different. A significant share of electricity is produced by Dnieper hydropower station, thus there is an opportunity of gradual transition to use of updating energy sources that do not threaten energy security. Therefore, the final decision about dismantling of hydropower stations should be made on the basis of comprehensive assessment of economic-ecological efficiency and evaluation of losses which can be caused by man-made tsunami.

We propose a complex approach to risk assessment of use of the Dnieper cascade hydropower station. We use a stochastic method of assessment of potential losses connected with the use of Dnieper reservoirs in order to assess the losses, which can be caused by violation of the integrity of the dam. We evaluated the potential losses of man-made tsunami for Kyiv reservoir. In the research was made evaluation of the potential hazards of each of the Dnieper reservoirs which can be caused by man-made tsunami. On the basis of the achieved results we ranked the reservoirs according to the degree of economic insecurity.

Transformation of the of the key symbol of the Ukrainian state of rapid flow into the system of stagnated reservoirs has no economic reasons taking into account that hydropower stations produce only 5% of the electricity of the total amount and the flooded areas can be used more efficiently. are more effectively use the flooded areas.

References

1. Veklych O. Ekologichna cina ekonomichnogo zrostannya Ukrainy. Ekonomika Ukrainy. 2012. 1. 51—60. (in Ukrainian)
2. Danylyshyn B. M., Dorogunczov S. I., Mishhenko V. S., Koval Ya. V., Novorotov O. S., Palamarchuk M. M. Pryrodno-resursnyj potencial stalogo rozvytku Ukrayiny. Kyiv: RVPS Ukrainy. 1999. (in Ukrainian)
3. Electronic resource Dnipro: http: uk.wikipedia.org (in Ukrainian)
4. Electronic resource of Derzhavnyj komitet statystyky: http://www.ukrstat.gov.ua/ (in Ukrainian)
5. Statystychnyj byuleten «Zhytlovyj fond Ukrayiny u 2013 roci» - Kyiv: Derzhavna sluzhba statystyky, 2014. S.8 (in Ukrainian)
6. Dnipro siogodni: tilky stogne, ale vzhe ne reve. Electronic resource Dzerkalo tyzhnia: http://gazeta.dt.ua (in Ukrainian)
7. Electronic resource Ystoryya Dneprogesa. Vzryv I vosstanovlenye: http://lifeglobe.net/ (in Ukrainian)
8. Electronic resource Kaskad Dniprovskyh vodosxovyshch: buty chy ne buty?: http://undiwep.com.ua/ (in Ukrainian)
9. Myxajlenko L.E., Lapshyn Yu. S., Vashhenko V.N. K voprosu o sostoyanya plotyn Kyevskoj GES. Derzhavna ekologichna akademiya pislyadyplomnoyi osvity ta upravlinnya. Naukovo-praktychnyj zhurnal «Ekologichni nauky» 2013, 2, 42-50. (in Ukrainian)
10. Electronic resource Ocinka zagroz gidrodynamichnoyi nebezpeky v Ukraini: http://ohranatrud-ua.ru/stati-po-gz/927-gidrodinamichnoji-nebezpeki-v-ukrajini.html (in Ukrainian)
11. PigouA. Ekonomycheskaya teoriya blagososto yaniya, Russia: English translation.-Moscow: Progress, 1985. (in Russian)
12. Coase Ronald The Problem of Social Cost, Journal of Law and Economics, 1960, 3(1), 1–44.
13. Maidment D.R. Handbook of Hidrology. NewYork.-1992.-Grow-Fill Inc.
14. Muller, Richard A. "Chapter 1. Energy, Power, and Explosions". Physics for Future Presidents, a text book. ISBN978-1426624599, 2001–2002
15. Samuelson, Paul A. “Diagrammatic Exposition of a Theory of Public Expenditure,” The Review of Economics and Statistics,1955, 37(4), 350–56.
16. J Andrew Charles, Paul Tedd, Alan Warren Delivering benefits through evidence, 2011.

Behavioral Aspects of Financial Anomalies in Ukraine

Tetiana Paientko

National University of State Tax Service of Ukraine
tpayentko@mail.ru

Abstract. This article is devoted to the problems of financial anomalies in Ukraine. Groups of main financial anomalies, and the key reasons for the development of such financial anomalies will be herein defined, and the behavior of the economic agents which frame financial anomalies in Ukraine will be explained. Possibilities for overcoming such financial anomalies will also be examined.

Keywords. Financial anomalies, economic behavior, revenue loss, shadow economy

Key terms. Model, Research

1. Introduction

The current state of the Ukrainian economy is most difficult. Government reforms which were decelerated have had a decisive impact on the further development of Ukraine as an independent nation. However, questions arise as to what the mechanism for implementing such reforms should be, and to the usefulness of implementing policies on the basis of foreign experience. The past few years show that most of the changes in the economy of Ukraine were as a result of taking into account foreign experience. However, applying such experience does not always result in the intended manner. One reason is that foreign policy examples were implemented quite imperfectly in the Ukrainian economy. Another reason is the underestimation of the time needed to properly implement reforms. Thirdly, one of the most significant reasons is the unexpected behavioral response of Ukrainian economic agents, which was quite different to reactions in other countries.

2. Theoretical and Methodological Background

Groundbreaking research work on understanding the fundamentals of the behavior of economic agents has been published by the leading scientists of institutional theory. In particular, D. North was one of the first who proved the existence of anomalies in an economy and finance that cannot be explained solely on the basis of economic laws (D. North, 1990).

J. Buchanan was one of the first who explained the role of social choice in the development of an economy and the reaction of economic agents on political decisions (J. Buchanan and G. Tullock, 1962). J. Stiglitz deeply investigated the causes of the global economic crisis of 2008, revealing the behavioral aspects and further consequences for a society (J. Stiglitz, 2011).

Research by Ukrainian scientists on the behavioral aspects of financial anomalies has been essentially unstructured. O. Pruts′ka explains how differences in the development of various societies are marked by reactions by members of the society to different types of externalities (O. Pruts'ka, 2003). A. Gritsenko has described aspects of economic anomalies in the economy of Ukraine and developed a classification of them (A. Gritsenko, 2003). V. Vishnevsky has researched the causes of financial anomalies (V. Vishnevsky, 2006). R. Pustovijt has investigated the nature of transaction costs in the economy of Ukraine (R. Pustovijt, 2004). Y. Ivanov and O. Jeskov maintain that one of the reasons for the failures of many reforms are attempts by the government to remedy the mistakes of the past without considering possible reactions by economic agents in the present (Y. Ivanov and O. Jeskov, 2007).

The causes and nature of financial anomalies in the economy of Ukraine have been studied using various methodological principles. Firstly, work was done on the basis of theoretical judgments and generalizations (O. Pruts'ka, 2003, A. Gritsenko, 2003), and secondly, using the tools of economics and mathematical modeling. Here arises another problem, because not all tools can be applied. For example, the use of correlation-regression analysis provides opportunity to describe the behavior of a group of agents (rather than the reaction of one agent) in specific terms for a specified period of time. This means that the use of predictive models based on data correlation for past periods could be incorrect. This explains the miscalculations in the development of state budgets, the failure of the planned indicators of budget, etc.

In this case, more accurate the results would be reached by fuzzy logic simulation modeling (V. Vishnevsky, 2006; O. Rajevnieva, 2007). However, in my opinion, such research should be complemented by the results of the application of game theory. It is a toolkit game theory and can explain the reasons for the behavior of each economic entity in a relevant situation.

3. Efficiency Estimation Procedure

The recent stage of development of the Ukrainian economy is characterized by a number of financial anomalies, which have been described in the publications of A. Grytsenko (2003), T. Paientko (2013), etc. Among the major financial anomalies are the following:

1. Deformed structure of economy of Ukraine, in which the dynamic development of the financial sector has not contributed to an increase in the volume of funding to the non-financial corporation sector. This issue is explored in detail in the article by T. Paientko and Y. Syrotiuk (2014).

The problem lies in the fact that the growth of the assets of financial institutions does not ensure the necessary growth of investments in the non-financial corporation sector of the economy. The crisis in Ukraine has further worsened the situation

regarding financing in that sector of the economy. First, bank lending has been actually paralyzed. The increases of the NBU discount rate initially to 19.5%, and then to 30%, have actually robbed banks of real opportunities to inject funds into the economy.

Secondly, in 2014-2015, 39 banks declared insolvency, and most banks have problems with liquidity and the ability to return deposits to customers. This was one of the causes of the bank panic and the outflow of deposits. The situation involving savers has worsened the steep inflation and devaluation.

Thirdly, problems with solvency have affected many insurance companies. This happened because they were placing their reserves mainly as deposits in banks, including those who have since become insolvent. Fourthly, mass poverty is developing within Ukraine, and that part of the population which forms the bulk of the depositors now appears on the brink of poverty.

Thus, the financial sector now finds itself on the brink of survival. The situation exists where the greater part of the population believes there is nowhere to invest. Furthermore, that part of the population that has savings in foreign currency will soon not be willing to inject funds into the financial sector. The behavior of economic agents in such situations can be described by using a toolkit of game theory. These are the possible strategies of a depositor and a bank:

1. The depositor puts money into a deposit account and the interest rate exceeds the rate of inflation (payout 1).
2. The depositor invests in a deposit account and the interest rate is lower than the level of inflation (payout 0).
3. The depositor puts money on deposit and the interest rate is lower than the inflation rate and the rate of devaluation (payout – 1).
4. The bank is ready to return the deposit by the end of the term together with interest (payout 1).
5. Temporary administration will be introduced in the bank during the term of the deposit. The depositor will receive compensation from the fund of guaranteeing deposits of individuals (payout 0).
6. Temporary administration will be introduced in the Bank during the term of the deposit. The depositor will not receive compensation (payout – 1).

Then the payout matrix will look this way (table 1):

Table 1. The matrix of payouts of the depositor and the bank

	Bank (4)	Bank (5)	Bank (6)
Depositor (1)	(1; 1)	(1;0)	(1; –1)
Depositor (2)	(0;1)	(0;0)	(0; –1)
Depositor (3)	(-1;1)	(-1;0)	(-1; –1)

N.B. – 1, 2, 3, 4, 5, 6 are strategies

As can be seen from the table, there is only one equilibrium strategy which provides a payout for both sides – (1; 1), which is possible with a probability of 1/9. Two strategies (0; 1) and (0; 0) do not provide payout for the depositor, with probability 2/9. The other strategies are without payout for the depositor with the

probability of 6/9. Potential depositors are unlikely to trust their savings to a bank because of this combination of circumstances.

There is a dilemma in such situations: the non-financial corporation sector of the economy requires an increase in funding, and the financial sector cannot provide it as a result of the outflow of funds. To overcome the described abnormalities, the government should take measures to stimulate the growth of personal savings. Reducing real income leads to a lower limit in the propensity to save. The drop in the propensity to save is now faster than the fall in real income.

2. Lack of correlation between the decrease in the tax burden and the dynamics of foreign investment in Ukraine's economy. This situation is also a financial anomaly caused by several institutional factors. Over the past twenty years, the Ukrainian government instituted significant tax benefits and other preferences for foreign investors. However, within the post-socialist space Ukraine remains an outsider in the attraction of foreign investments per capita. In addition, most foreign investment is coming into Ukraine from regions where there exists a more favorable investment climate, offshore entities, and Russia (table 2).

Table 2. Foreign direct investment (equity) in Ukraine's economy,%

Indicators	2010	2011	2012	2013	2014
Total	**100,0**	**100,0**	**100,0**	**100**	**100,0**
Which includes					
Cyprus	22,2	25,6	31,7	32,7	29,9
Germany	15,8	15,0	11,6	10,8	12,5
Netherlands	10,5	9,8	9,5	9,6	11,1
Russia	7,6	7,3	7,0	7,4	5,9
Austria	5,9	6,9	6,2	5,6	5,5
United Kingdom	5,3	5,1	4,7	4,7	4,7
Virginia Islands (Brit.)	5,1	4,5	3,5	4,3	4,4
France	3,9	3,5	3,2	3,1	3,5
Switzerland	3,3	3,3	2,9	2,3	3,0
Italy	2,7	2,1	2,0	2,2	2,2
USA	2,2	2,0	1,9	1,8	1,9
Poland	2,1	1,9	1,7	1,7	1,8
Belize	1,9	1,8	1,7	1,5	1,4
Other	11,5	11,2	12,4	12,3	12,2

As is evident from the data presented in table 2, the largest volume of foreign investment in Ukraine’s economy is coming from Cyprus. In its essence it is not an investment, but the return of capital removed previously from Ukraine. In most developed economies, providing tax incentives to foreign investors provides an increase in foreign investment. In Ukraine, this tool does not work. According to the World Investment Report and Ranking, the reduction of business taxation is not a determining factor when deciding on investing in Ukraine. Even before the beginning

of the armed conflict, key analysts and potential investors indicated greater concern over the issues of the low level of protection of property rights and the high level of corruption. Domestic investors are also not actively investing in the domestic economy. On the contrary, much of the internal capital has been removed from Ukraine. This is an extra negative indicator for foreign investors.

Investors (domestic or foreign) make investment decisions taking into account the following probabilities:

1. The government will change the rules of the game and preferences for foreign investors will be eliminated – p (A).
2. The prevalence of bribing – q (B).
3. The infringement of ownership rights of the investor – $1-(p+q)$ (C).

Probable scenarios of the government can be described as follows:

1. An investor makes a decision about investing in the Ukrainian economy in spite of the existing risks. Investments are long-term. This is an absolute win for the economy, which denotes 2 (if we assume that 0 is the loss to the state in the absence of investment).
2. The investor does not assume all of the risk, but decides to invest in the economy. However, such investment is generally directed into short term projects intended for a fast return. Under such circumstances the economy would win, but it is smaller than the previous version – 1 (B).
3. The investor takes no risks and decides not to invest in these conditions (C).

The described version is a game that will repeat. There is the possibility that a future investor will change his course of action. However, the probability of investor choices changing depends on how the government shapes the business environment. The payoff matrix is presented in table 3.

Table 3. Game Matrix: Investor and the Government

The choice of the investor	The choice of the Government: A	B	C
A	(1;2)	(1; 2)	(–2; 2)
B	(1; 1)	(1;1)	(0; –1)
C	(0; 0)	(0; –1)	(0; –2)

Source: compiled by author

In this case the probability of investor payout can be described as:

The probability in a change of rules of the game initiated by the government:

$p+q+0(1-(p+q)= p+q$

The probability of a bribe being requested:

$p+2q-1\ (1-(p+q)) = 2p+3q-1$

The probability of infringement on ownership rights:

$2p-q+0(1-(p+q) = 2p-q$

All three options can be equally acceptable for domestic investors. As in the case of limited foreign investment the cost of domestic investment increases. However, this must be true:

$p+q = 2p+3q-1 = 2p-q$

Having solved the equation, we obtain: q = ¼, p = ½, (1 – p – q) = ¼

So, the most decisive factor for investors is a change of the rules of the game by the government. They believe that this risk is the largest. However, you can see that the risk of bribing or infringement of rights is smaller. The risks do not stand alone. Their real impact is expressed only with the risk of changes to the rules of the game. It means that q = (1 – p – q) =¾. Under such conditions the likelihood of foreign investment is preserved. However, it would likely be short-term investments in projects with a fast turn-around period. Therefore, an improvement in the investment climate in Ukraine provides for a stabilization of the rules of the game for the investor. The investor should be guaranteed that the rules of the game would not change within a fixed period of time.

3. Lack of correlation between the size of the tax burden and the dynamics of the informal sector of the economy. One of the greatest problems for the Ukrainian economy is the degree of its shadow economy. According to various experts the volume of the shadow economy in Ukraine constitutes 40% (according to the Ministry of Economic Development and Trade of Ukraine) to 80% (estimated by the Schneider Institute). There is a misconception that the shadow sector of the economy in Ukraine was formed after the breakup of the Soviet Union. However, shadow economic activity existed in the times of the USSR.

The policy of "war communism" (1918-1921), from the outset, carried within elements of shady dealings. It meant that the government resorted to violent methods and centralized administrative pressures to accomplish their own goals. In those times speculation, gangsterism, and robbery developed rapidly (R. Viseberg, 1925, p. 43). In the time of Stalin about 30% of production was embezzled from socialist enterprises, and about a quarter of the resources redistributed centrally was not by intention.

A sharp reduction in the non-government sector in the late 1950s – early 1960s, contributed to the further development of the informal sector of the economy. Commercial cooperatives were eliminated in that period, the final transition from state farms to collective farms happened, prohibitions of individual trade restrictions on keeping personal subsidiary plots were decelerated, as was the ban on the holding of cattle, etc. A trend towards further consolidation of production, an increasing phenomena of monopolies in economics, ideological mandates proclaiming a further transition to communism - all these factors shaped an economy with dual sectors − official and shadow, which interacted with one another. Thus began the emergence of speculative markets, clandestine workshops, black marketeers and speculators in foreign currency. According to modern estimates, the 1960s saw that unofficial production supplied 20% of industrial products, 40% of food products, and about 35-45% of all scarce consumer goods was made available through speculative markets.

During the era of Brezhnev, the shadow economy flourished and took almost an official color. By the beginning of the 1980s, all regions of the country began to encounter clandestine workshops, using the state's equipment, material, and energy resources, and funds. Production surpassed mandated limits which were set by the State Supply and State Planning Commission, and the income was widely distributed (T. Koriagina, 1990).

This shadow economy, tolerated by Leonid Brezhnev, yielded by the most generous estimates 10-15% of GDP, and then rose to 50% of GDP in the period before "Perestroyka". The level of corruption between 1980-1985 in the Soviet Union put it in the middle of a ranking of 54 countries, having a larger bureaucracy than Italy, Greece, Portugal, South Korea and virtually all developing countries.

In the USSR at the beginning of the 1990s, the volume of the shadow sector was assessed to be in the amount of 100 billion rubles by average valuation, 20-25 billion rubles from the most conservative, and some pegged it at 150 billion rubles. In comparison with the beginning of the 1960s, the growth of the scale of the shadow economy across the whole range of ratings was from 4 to 30 times (T. Koriagina, 1990).

With the collapse of the USSR the Soviet shadow economy ended. However, shadow economies began to resurface in the individual independent republics, and the specific conditions and trends within each new country determined the dynamics and scope of illegal operations. According to various estimates the volume of the Ukrainian shadow economy at the beginning of its existence as an independent state (1991) was estimated at 18% of GDP. Thus, the Government had from the outset made an error in believing the tax burden was the main factor in the development of the shadow economy in Ukraine. It ignored other contributing factors for the development of a reducing shadow economy. Therefore, the reaction of the economic agents was not as expected by the government.

After the collapse of the USSR the shadow sector continued to grow. A decisive role in this was played not only by irrational tax policy. The development of shadow economic activity contributed to hyperinflation, an increase in bartering, and the opportunistic behavior of civil servants, etc. The growth of the tax burden in the period 1991-1996 also played a negative role. However, we are not merely noting the direct interdependence between the sum of money required to pay for the benefit of the government, and the amount of shadow activities. A principle motive of shadow economic relations and tax evasion was that taxpayers did not trust the government to responsibly use tax revenues to the benefit of the greater society. This type of economic agent truly believed that they could better use the savings than had the government had the funds.

For an explanation of the behavior of economic agents, it is advisable to use a model of expected utility, which was developed by Von Neumann and Morgenstern (Von Neumann and Morgenstern, 1970). According to the standard model of choice in conditions of uncertainty for taxpayers, it is also a game. They estimate the payment of taxes in accordance with the expected utility. In conditions of uncertainty, such behavior is described by the prospects theory. In general this theory is as follows: Assume, the taxpayer has to play a lottery (x, p, q; y). This means that the lottery has a result x with probability p and the result y with a probability of q. The taxpayer assesses this lottery this way (1):

$$\pi(p)\, v(x) + \pi\,(q)\, v\,(y) \qquad (1)$$

v(x) – function values that the individual gives the winning or losing and $\pi(p)$ – weight, which the individual provides objective probabilities when making decisions. Hypothetical functions v and p are presented on fig. 1.

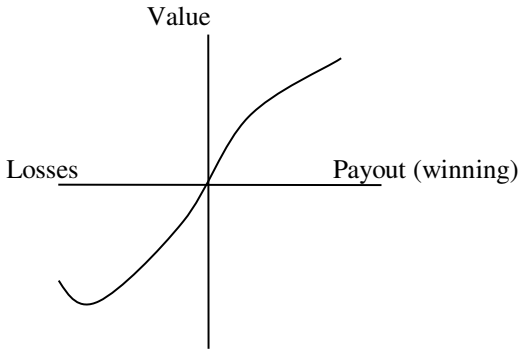


Fig. 1. Hypothetical function values

This theory has three important features. First, the value of winnings and losses are defined separately. This is consistent with the analysis of games and the choice of people in conditions of risk. Second, it is a form of function values. Function is concave in the interval of winnings and convex in the range of losses. This means that taxpayers are shunning risk in the winning area and attaching risk to the losses area. The function values must be 0 at the point of starting. It means there is more extreme sensitivity to losses than to gains. This trait is called avoidance of losses. The third feature is the nonlinear transformation of probability π (p). Unlike probability p, π (p) is a weight that provides an objective probability when making decisions. As a rule π (p) = p, but π (p)<p for large p. Small probability receives a relatively large weight, with π (p)>p (A. Lukashov, 2004, pp. 40-41).

Function definition of weight is also characterized by subcertainty property: for all $0<p<1$, $\pi(p)+\pi(1-p)<1$. The principle of incomplete probability describes the attitude of people to probable events. The weight of the two probabilities of complementary events is less than the weight of the one event that must occur with a probability of 100%.

Taxpayers are more sensitive to the difference in probability at its higher levels. According to experimental data, taxpayers would surely want to hide 10,000 UAH in taxes at 100% certainty than receive social benefits from the state totaling 12,000 UAH with a probability of 80% that they would appear, while taxpayers believe that it is better to have 12,000 UAH with a 20% probability, than 10,000 UAH with a probability of 25%. According to prospects theory:

$$\frac{\pi\,(0{,}25)}{\pi\,(0{,}2)} = \frac{\pi\,(1)}{\pi\,(0{,}8)} \qquad (2)$$

So, a 20 point increase in the probability of 0.8 to 1.0 has a greater effect than an increase from 0.2 to 0.25. Therefore, lowering the tax rate on corporate income in 2004 (from 30 to 25%) did not cause a reduction of the shadow economy since the probability of the growth of volumes of goods from the government was less than the amount of unpaid taxes. Lowering the tax rate on corporate income between 2010 to 2014 (from 25% to 18%) also did not contribute to a reduction of the shadow economy. Since the weight of probability of winning from the non-payment of taxes

exceeds the probability of a return of government benefits. Based on the foregoing, the key factor in the reduction of the informal sector of the economy is changing the behavior of the government, not the reduction of the tax burden.

Taking into account the results of research of previous financial anomalies, one can expect businesses to exit from the shadows if the following conditions can be fulfilled:

1. A change in the behavior of the government. Taxpayers must trust how taxes are being utilized.
2. Preservation of property rights should be guaranteed and not merely proclaimed.

The next financial anomaly is closely associated with the above-described situation. It is possible to explain how to use game theory (a zero sum game) by using the model of expected utility and prospects approach. However, the best model in this case would be the principal-agent theory.

4. Lack of meaningful communication between established punitive sanctions for violations of tax legislation and the level of taxes.

Low taxes are the problem, which every Government in Ukraine is trying to overcome. To improve the level of payments of taxes administrative methods were mainly applied. In particular, there was an increase in the number of grounds for carrying out unscheduled inspections and increasing the size of penalties. However, this failed to achieve the desired level of tax payments. Also, the activities of tax officials is characterized by low efficiency (tabl. 4).

Table 4. The results of the activity of tax police in Ukraine for 2007-2013, million UAH

Indicators	2007	2008	2009	2010	2011	2012	2013
Extra revenue for budget, discovered by tax police	259.94	382.1	502.68	4370.59	1052.2	2082.76	2014.5
Charge involving tax police	130.39	173.53	3461.81	451.67	568.67	978.6	1023.2
The amount stipulated damages in criminal cases of tax evasion	2429.78	2662.84	2406.57	2008.86	2173.89	1997.6	2117.4
Refunded the sum of damages in criminal cases of tax evasion	878.30	1155.97	829.52	816.78	869.71	888.2	902.3

As can be seen from table 4, even in those years when there was a growth in the shadow economy (2012-2013), large increases of revenues to the budget from punitive penalties did not occur. The tax police did not execute one of their main functions – to provide reimbursement of losses by the government.

Throughout 1997-2015, the government tried to increase the size of penalties for violations of tax legislation several times. However, the discipline of taxpayers has not changed. The approach used by regulatory agencies prior to the imposition of penalties has not changed. Each year, the regulatory agencies set a planned amount of punitive penalties. This means that the same amount of fines would have to be

recovered from taxpayers by the regulators. If the amount was less, then the head of the relevant local authority would have to explain why the targets had not been reached. If the amount collected was greater, then the following year the planned penalty targets would be increased.

Such an approach was borrowed from Soviet times. It is false from the very beginning. It thwarted an opportunity to build partnerships between the government and the taxpayer. It violates clause 4.1.4 Article 4 of the tax code of Ukraine about a presumption of legality. Also, the actions of regulatory authorities often violate clause 4.1.2 c. 4 of the tax code of Ukraine on the equality of all taxpayers. In practice, it has not been uncommon to have cases of selective application of penalties. Such cases discredit the image of the government and are not conducive to an increase in the level of trust.

The question arises as to why these facts have a place. Indeed, at first glance, this behavior is aberrant. This anomaly can be explained in terms of the theory of agency relations. In this situation, the government is the principle. It sets the rules of the game, based on existing information. However, the information is incomplete. Regulators and taxpayers are agents who in real life have more information on a specific situation.

It is a situation where both agents, if they want to be profitable, must behave opportunistically. The taxpayer knows that in any case he will have to pay a penalty (a plan of fines). The tax inspector must collect a minimum amount of fines. They are not interested in overfulfillment of the plan. It is easier for taxpayers and a tax inspector to engage in conspiracy and agree on the amount of the penalty. As a result, the taxpayer may violate tax law and not expect a higher responsibility than that agreed. Instead, the inspector receives a bribe and there is a loss of appropriate fines and charges. There is a fairly simple way out of this situation. The government should cancel the plan for punitive sanctions and reduce the number of cases of direct communication of the taxpayer and the controller.

4. Conclusions

The described financial anomalies have a serious negative impact on the potential for economic development. In terms of the theory of behavioral finance, the situation is described as abnormal. They are the behavior reactions of economic agents to the challenges of the environment. The economic policy of the government should consider not only the potential economic effects, but also the expected behavioral response of economic agents. During the development of the economic policy of the government must pay attention primarily to such behavioral aspects:

1. Economic agents make decisions based on available information. Therefore, information about the real economic situation should be fully disclosed. This will reduce the information asymmetry in the relationship of the government to the economic agent.

2. Economic agents make decisions based on the maximization of value, utility, and the expected probability of receiving benefits. Therefore, the government should ensure the provision of quality public benefits on the basis of tax revenues.

3. Economic agents invest in the economy if there is a guarantee of preservation of property rights. For the investor, the decisive factor is that the rules of the game remain constant from the government during the term of the investment. Therefore, the government must ensure the stability of conditions for investment over long-term periods.

In essence, consider behavioral aspects when developing economic policies to overcome existing financial anomalies and to avoid the emergence of new ones.

References

1. North D.: Institutions, Institutional Change and Economic Performance, Cambridge University Press (1990).
2. Buchanan J.M., Tullock G..: The Calculus of Consent: Logical Fundations of Constitutional Democracy. Ann Arbor: University of Michigan Press (1962).
3. Stiglits J.: America, Free Markets, and the Sinking of the World Economy. Moscow, EKSMO (2011).
4. Prutska O.: Institutionalism and Problems of Economic Behavior in Transition Economy. Kyiv, Logos (2003).
5. Grytsenko A.: Features of the Institutional Structure of Ukrainian Society in the 21st Century. Ukraine's Economy: Strategy and Long-Term Development Policy. Kyiv, Institute of Economics and Forecasting, Phoenix (2003).
6. Vyshnevsky V., Vetkin A., Vyshnevskaya E.: Taxation: Theories, Problems, Solutions. Donetsk, Donetsk IEP (2006).
7. Pustoviyt R.: Transaction Costs: Theoretical Concepts and Empirical Analysis. Economist 10, 26–29 (2004).
8. Ivanov Yu., Yeskov O.: Modern Taxation: the Motivational Aspect. Kharkiv, INJEK (2007).
9. Raievneva O., Goliad N.: Simulation of Anti-Crisis Management of Region. Kharkiv, INJEK (2007).
10. Paientko T.: Institutionalization of Fiscal Regulation of Financial Flows. Kyiv, DKS center (2013).
11. Paientko T., Syrotiuk Yu.: Accumulating of Financial Resources by Financial Intermediaries and its Influence on Economic Development. Business infom. 8, 237–243 (2014).
12. Vaisberg P.: Money and Prices (An Underground Market in the Period of "War Communism"). Moscow, State plan publishing (1925).
13. Koriagina T.: The Shadow Economy in the USSR. Questions of economy 3, 29–41 (1990).
14. Von Neiman J., Morgershten O.: Theory of Games and Economic Behavior. Moscow, Science, (1970).
15. Lukashov A.: Behavioral Corporate Finance and the Company's Dividend Policy. Management of Corporate Finance 2, 35–47 (2004).

The Formation of the Deposit Portfolio in Macroeconomic Instability

Andriy Skrypnyk[1], Maryna Nehrey[1]

[1] National University of Life and Environmental Sciences of Ukraine
avskripnik@ukr.net, Marina.Nehrey@gmail.com

Abstract. In 2014 the main tendency of Ukrainian economy was the losing of great deposit value. In this article we wish to explore a deposit portfolio structure in macroeconomic instability. We applied two approaches to the standard optimization portfolio: risk minimization for a given maximum return and return maximization for a given maximum risk. Of the two approaches to the standard optimization problem of portfolio: risk minimization at a given minimum return and return maximization for a given maximum risk the advantage was given the latter. The exchange rate risks are the main factors that have a significant impact on the end result. The optimum structures deposit portfolio was calculated for six different situations in national and world financial markets. Comparison of the optimal portfolio structure with real historical data showed that customers of the banking system over evaluate the reliability of the financial system.

Keywords. deposit, devaluation, portfolio, optimization, return, revaluation, risk.

Key Terms. Data, DecisionSupport, Development, FormalMethod, Management, MathematicalModel.

1 Introduction

The unstable macroeconomic situation in Ukraine and the crisis of the banking system caused distrust in the banking institutions. According to the opinion of experts, the Ukrainian population kept at home cash equivalent to $10 billion USA. In recent years was observed the following tendency: in 2014 banks lost deposits in the amount of 126 billion UAH, and around 18 billion UAH during first two months of the current year [3]. However, storage of money at home has several disadvantages: for example lack of income from capital and high risks, which lead to additional costs for the implementation of the safety of their own homes and significantly decrease the level of living.

Banking experts usually advise to divide money into three equal parts, two of which are nominated into euros and US dollars according to the current exchange rate, and put on deposit accounts in different banks which can be considered reliable

(it is advisable to choose banks which are included in the deposit insurance program NBU) and wait for interests during this period (simple diversification). Unfortunately, this method is connected with difficulties. It is almost impossibile to convert legally the accumulated funds into any reliable currency, besides it is rather difficult to find a reliable bank. This study is limited to two currencies - US dollars and euros, however, presented method can be used to form a deposit portfolio using other currencies.

There are two approaches to the portfolio optimization problem: risk minimization at a given minimum return and return maximization for a given maximum risk. For portfolio optimization you need to determine in which currency to evaluate the result. We can ask a question: "Why do we save money?" The answer can be the following: "In order to increase consumption during our life (real estate, household appliances, automobiles, traveling)" [2]. The vast majority of consumed goods in Ukraine are produced outside the country and therefore it is better to measure the cost by the most stable currency, which is now can be considered the US dollar. Alan Greenspan devoted attention to keeping a low dollar inflation level than in the past since such a policy, combined with the larger predictability of monetary policy, contributed to making dollar capital denomination most attractive [11].

2 Markowitz Problem under Devaluation Condition

The Markowitz's portfolio optimization problem can be solved using the well-known term of return and risk (variance of return) components portfolio. If return is measured as the deposit interest, the rate of risk is measured by its dispersion [4]. Linear model was proposed for credit risks in order to maximize bank profit [6, 10]. However, there is a factor that has a significant impact on the end result - an exchange rate risks, which is more important for unstable economics [3]. Of course interests on deposit and credit accounts for exchange rate risks, as the interest on UAH deposit twice as much than the dollar deposit [1, 12]. The importance of foreign exchange component in the sustainability of the banking system was emphasized in a number of research [5, 13]. In this study we wish to evaluate the optimal structure of the deposit portfolio during economic turbulence and make a comparison between real and optimal structure deposit portfolio.

Exchange rate risks can be taken into account, if a devaluation matrix is specified.

We will consider the case-study of placing deposits for one year. We assume that three macroeconomic situations, which determine the devaluation processes in the country $\theta_1; \theta_2; \theta_3$, which are defined probabilities $p_1; p_2; p_3 (\sum_{i=1}^{3} p_i = 1)$. Each situation corresponds to a certain devaluation factor relative to USD defined as the ratio of the exchange rate in a current moment to exchange rate what will be in a year. We will denote devaluation multiplier for each economic situations $\phi_i (i = 1,2,3)$. If we know the value of a random variable and the corresponding probabilities, we can estimate the expected value of depreciation factor and its variance:

$$\bar{\phi} = \sum_{i=1}^{3} p_i \phi_i ; \sigma_{\phi}^2 = \sum_{i=1}^{3} p_i \phi_i^2 - \bar{\phi}^2 \qquad (1)$$

Later we will consider the case of uniform distribution of devaluation multiplier.

If $\bar{\phi} < 1$ then dominate devaluation expectations, if $\bar{\phi} > 1$ then dominate revaluation expectations. There were short periods of revaluation of UAH, but we observe the tendency of devaluation according to results of any year.

It is supposed to use the share denominated in euro for deposit portfolio, which has currency instability relative to leading world currencies and the objective function is denominated in USD, we need to specify the expected devaluation and its variance in EUR against the USD for the next year. We will denote these parameters: $\bar{\varphi}; \sigma_{\varphi}^2$.

In this formulation dollar deposits is completely risk-free, which is rather optimistic assumption. During the year, the interest on dollar deposits was changeable, which can be used as a risk assessment. We denote the variance of interests on USD deposits $\sigma_{\$}^2$. We assume that the current interest on USD deposits is in the interval 9-11% [8] and is characterized by a uniform distribution, the dispersion interest is approximately equal $\sigma_{\$}^2 \approx 3{,}3 \cdot 10^{-5}$.

We consider the standard formulation of the Markowitz problem taking into account the expected devaluation (revaluation) processes.

We present the particles deposit portfolio in UAH, EUR and USD: $d_1; d_2; d_3$ $(d_1 + d_2 + d_3 = 1)$, percentage interests $r_1; r_2; r_3$ $(r_1 >> r_2 > r_3)$ are ranged under level of risk of deflationary expectations. If an initial investment is S_t than in a year the expected amount of the deposit portfolio and its dispersion will be:

$$\begin{aligned} S_{t+1} &= d_1 S_t (1 + r_1)\bar{\phi} + d_2 S_t (1 + r_2)\bar{\varphi} + d_3 S_t (1 + r_3), \\ \sigma_{\Pi}^2 &= d_1^2 S_t^2 \sigma_{\phi}^2 + d_2^2 S_t^2 \sigma_{\varphi}^2 + d_3^2 S_t^2 \sigma_{\$}^2. \end{aligned} \qquad (2)$$

There are no members in portfolio variance that appear as a result of presence of the connection between return components of portfolio. The reason is that in this case independent devaluation processes influence on the profitability: euro and US dollar and the processes of devaluation of the national currency because of macroeconomic instability in the country. Therefore, we can assert absence of connection between return of the portfolio shares denominated in different currencies in the proposed formulation.

If the level of devaluation is high, the depositor will have loses $(S_{t+1} < S_t)$, that is why we will limit the possible risk-free profit according to the interest which is equal to r_3 (the return of dollar deposits):

$$d_1 S_t (1 + r_1)\bar{\phi} + d_2 S_t (1 + r_2)\bar{\varphi} + d_3 S_t (1 + r_3) > S_t (1 + r_3) \qquad (3)$$

From the last expression we can get maximum portfolio share of deposits denominated in UAH $\bar{\varphi} = 1$:

$$d_1 < \frac{d_2(r_2 - r_3)}{1 + r_3 - (1 + r_1)\bar{\phi}} \tag{4}$$

We estimate the maximum share of UAH deposits in terms of catastrophic devaluation in 2014. The difference in interests denominated in euros and dollars is less than 2%, the maximum value of the numerator is less than 0.01.

Devaluation multiplier for the previous year is approximately equal to 0.4 (8 USD / UAH 20 = 0.4). Interests on deposits are $r_1 = 25\%; r_3 = 10\%$. Therefore, the share of UAH deposits in terms of landslide devaluation should not exceed 2%.

3 Optimal Portfolio Structure

We estimate the portfolio structure with maximum profitability and limited risks for different combinations of UAH/USD and EUR/USD devaluation multiplier factors. Evaluation of devaluation multiplier factors is based on monthly time series of UAH/USD (03.1997 - 02.2015) and EUR/USD (02.2007 - 02.2015) exchange rates.

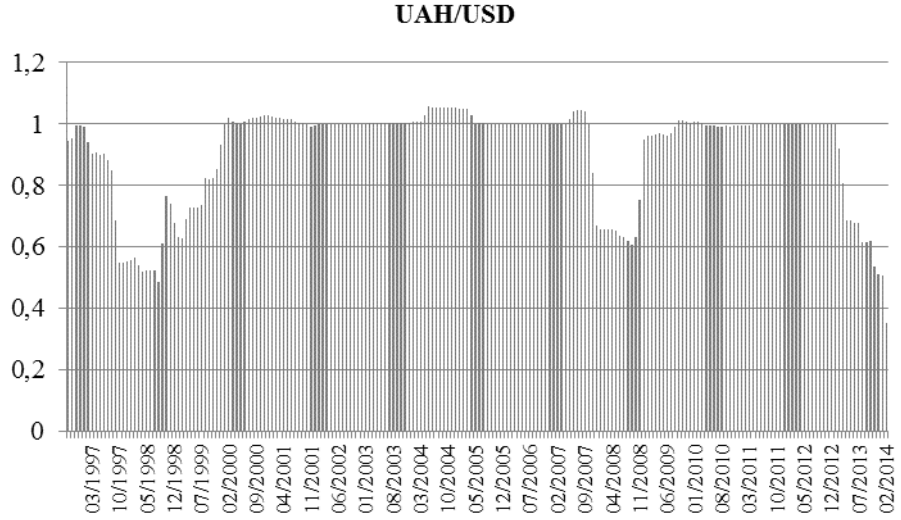


Fig. 1. Dynamics of devaluation multiplier UAH/USD

Devaluation multiplier measured with one year interval (deposit time in optimization problem) and currency pairs we calculated every month from March 1997 to February 2014 (210 observations UAH/USD) and form February 2007 to February 2014 (98 observations EUR/USD). (Fig. 1, 2).

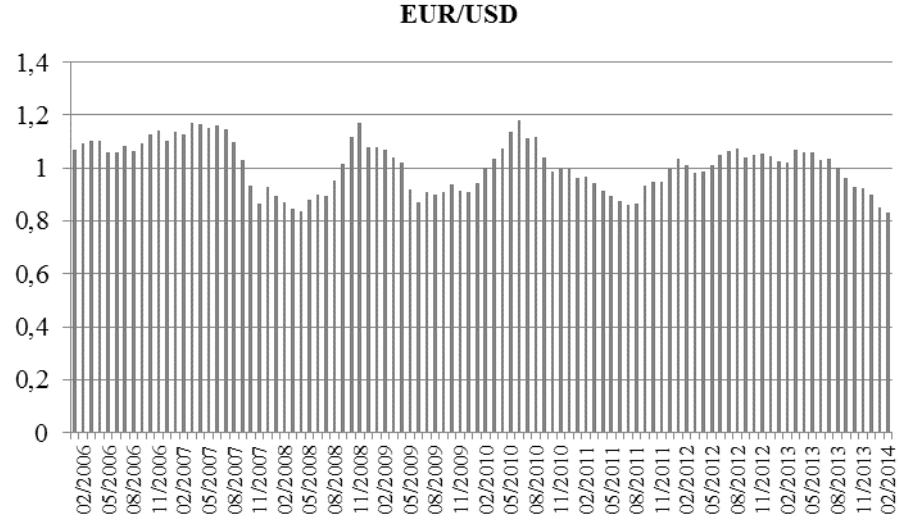


Fig. 2. Dynamics of devaluation multiplier EUR/USD

The period (1997-2014 for UAH/USD) consists of periods of economic growth with fixed course and periods crisis when monetary system tends to new equilibrium.

Devaluation multiplier factor UAH/USD $\phi \geq 1$ under 155 observations (minor revaluation probability $p_r = 0{,}736$), $\phi < 1$ under 55 observations (devaluation probability $p_d = 0{,}264$). Devaluation multiplier factor EUR/USD $\varphi \geq 1$ under 44 observations (revaluation probability $p_r = 0{,}449$), $\varphi < 1$ under 54 observations (devaluation probability $p_d = 0{,}551$).

Devaluation multiplier EUR/USD has more natural character, when the equilibrium is set under the influence of many non-interrelated reasons and a stable tendency is missing. The stationary hypothesis of the exchange rate of EUR/USD can be proved if we explored a long time period. The same hypothesis for exchange rate of UAH/USD must be rejected because of a full asymmetry of devaluation multiplier relatively to unity level.

We consider the optimal portfolio structure in three cases: landslide devaluation from 43% to 150% - θ_1 *(* $\phi \leq 0{,}7$ *);* moderate devaluation of 11% to 43% - θ_2 *(* $0{,}7 < \phi \leq 0{,}9$ *);* and a devaluation less than 11% - θ_3 ($0{,}9 < \phi \leq 1{,}0$). We regard the distribution of devaluation multiplier at each of the intervals being uniform.

We consider two possible states in the global financial market for devaluation multiplier for EUR/USD: θ_1^C *(* $0{,}8 \leq \varphi < 1{,}0$ *)* and revaluation multiplier: θ_2^C *(* $1{,}0 \leq \varphi \leq 1{,}2$ *)*. We present six possible situations that correspond to two situations of the world finance market (the euro-dollar) and three situations of devaluation in the domestic market (Table 1).

We have used interests of one-year deposits in banks of first group (the most reliable) to build optimization models. Of course, other banks interests can be significantly higher, but in this case it is necessary to increase the risk measures of bankruptcy probability due to the growth (receiving contributions under the insurance program of NBU connected with the loss of time and interest and primary contribution for more than 200 thousands UAH). We use the current annual deposit interests February 2015: $r_1 = r_U = 23\%; r_2 = r_E = 13\%; r_3 = r_\$ = 12\%.$

Table 1. Expected value devaluation factors for different classes of national and world economies in 2015

	θ_1 ($\phi \le 0{,}7$);	θ_2 ($0{,}7 < \phi \le 0{,}9$);	θ_3 ($0{,}9 < \phi \le 1{,}0$)
θ_1^C ($0{,}8 \le \varphi < 1{,}0$)	$\bar{\phi} = 0{,}55; \sigma_\phi^2 = 7{,}5 \cdot 10^{-3}$ $\bar{\varphi} = 0{,}9; \sigma_\varphi^2 = 3{,}3 \cdot 10^{-3}$	$\bar{\phi} = 0{,}8; \sigma_\phi^2 = 3{,}3 \cdot 10^{-3}$ $\bar{\varphi} = 0{,}9; \sigma_\varphi^2 = 3{,}3 \cdot 10^{-3}$	$\bar{\phi} = 0{,}95; \sigma_\phi^2 = 0{,}8 \cdot 10^{-3}$ $\bar{\varphi} = 0{,}9; \sigma_\varphi^2 = 3{,}3 \cdot 10^{-3}$
θ_2^C ($1{,}0 \le \varphi \le 1{,}2$)	$\bar{\phi} = 0{,}55; \sigma_\phi^2 = 7{,}5 \cdot 10^{-3}$ $\bar{\varphi} = 1{,}05; \sigma_\varphi^2 = 0{,}8 \cdot 10^{-3}$	$\bar{\phi} = 0{,}8; \sigma_\phi^2 = 3{,}3 \cdot 10^{-3}$ $\bar{\varphi} = 1{,}05; \sigma_\varphi^2 = 0{,}8 \cdot 10^{-3}$	$\bar{\phi} = 0{,}95; \sigma_\phi^2 = 0{,}8 \cdot 10^{-3}$ $\bar{\varphi} = 1{,}05; \sigma_\varphi^2 = 0{,}8 \cdot 10^{-3}$

We consider the problem of calculation of the share of certain currencies in deposit portfolio that maximizes the return of the portfolio for a given maximum risk level, which is equal to variance of interests on USD deposits:

$$\sigma_\$^2 = \frac{(r_\$^{\max} - r_\$^{\min})}{12}. \tag{5}$$

For $r_\$^{\max} - r_\$^{\min} = 0{,}02 \;\; \sigma_\$^2 \approx 3{,}3 \cdot 10^{-5}$.

We obtain the following problem to be resolved for finding d, $d = (d_1; d_2; d_3)$:

$$\begin{aligned} & S_{t+1} = d_1 S_t (1 + r_1)\bar{\phi} + d_2 S_t (1 + r_2)\bar{\varphi} + d_3 S_t (1 + r_3) \to max \\ & d_1^2 S_t^2 \sigma_\phi^2 + d_2^2 S_t^2 \sigma_\varphi^2 + d_3^2 S_t^2 \sigma_\$^2 \le \sigma_\$^2, \\ & d_1 < \frac{d_2 (r_2 - r_3)}{1 + r_3 - (1 + r_1)\bar{\phi}}, \\ & \sum_{j=1}^{n} d_j = 1, \\ & d_j \ge 0, \; j = \overline{1,3}. \end{aligned} \tag{6}$$

We analyze the results of the calculation of the structure of deposit portfolio with maximum return, depending on the situation in the global and domestic foreign currency markets (Table 2).

There are six situations according to the number of components in Table 2: (1, 1) - moderate devaluation of the euro and the significant UAH depreciation; (1, 2) - moderate devaluation of the euro and the moderate devaluation of the UAH (1, 3) - moderate devaluation of the euro and slight currency depreciation; (2, 1) - moderate appreciation of the euro and the significant currency depreciation; (2, 2) - moderate appreciation of the euro and moderate currency depreciation; (2, 3) - moderate appreciation of the euro and the slight depreciation of the UAH.

Table 2. Optimization of deposit portfolio according to the criterion of profit maximization

	θ_1 ($\phi \le 0{,}7$);	θ_2 ($0{,}7 < \phi \le 0{,}9$);	θ_3 ($0{,}9 < \phi \le 1{,}0$)
θ_1^C ($0{,}8 \le \varphi < 1{,}0$)	$d = (0;0;1)$ $S_{t+1} = 1{,}12$	$d = (0;0;1)$ $S_{t+1} = 1{,}12$	$d = (1;0;0)$ $S_{t+1} = 1{,}1685$
θ_2^C ($1{,}0 \le \varphi \le 1{,}2$)	$d = (0;0{,}73;0{,}27)$ $S_{t+1} = 1{,}1685$	$d = (0;0{,}73;0{,}27)$ $S_{t+1} = 1{,}1685$	$d = (0;1;0)$ $S_{t+1} = 1{,}1865$

In cases (1, 1) and (1, 2) optimal portfolio contains only dollar deposits with certain return. In the case (1, 3) portfolio consists only of UAH deposits (the return is corrected to the expected depreciation up to 11.1%).

In cases (2, 1) and (2, 2) the same return is defined by 73% share of deposits nominated in euros and 27% of deposits nominated in dollars. In the case (2, 3) the return which is equal to 18.65% is defined by 100% share of euro deposit. However, it is better to based the assumptions on mathematical forecast about the structure of portfolio that depends on the probabilities of the external environment: p_i – the probability of devaluation i state (i = 1,2, ..., k) cross currency exchange rate UAH/USD, q_j – the probability of the depreciation of the j-th state (j = 1,2, ..., n) cross currency exchange rate EUR/USD, $p_{ij} = p_{i\cdot} \cdot q_j$ – the probability of simultaneous occurrence of the i and j devaluation states, d_{ij} – the optimal portfolio structure according to i devaluation state of the UAH/USD and j state pair EUR/USD. Expected portfolio structure is defined as:

$$\overline{d} = \sum_{i=1}^{k} \sum_{j=1}^{n} p_{ij} d_{ij} \ . \tag{7}$$

We calculate the expected portfolio structure, assuming that the devaluation and revaluation expectations of the euro-dollar are equal.

($p_1 = p_2 = 0{,}5$), the first basic variant is calculated according to the assumption that all three devaluation states have the same devaluation probability (it is a situation

of absolute uncertainty). That is why $p_{ij} = 1/6$. This is basic structure of the portfolio and its expected return:

$$\bar{d}_{Б} = (0{,}167;0{,}41;0{,}423)\ldots \bar{r}_{Б} = 15{,}53\%; \sigma^2_{Б} = 7{,}4 \cdot 10^{-4}.$$

We consider pessimistic option in which the probability of a significant devaluation is twice higher than the probability of low, moderate devaluation and probabilities moderate devaluation is equal to the sum of probabilities of large and small devaluation:

$$p_{ij} = \begin{pmatrix} 2/12 \ldots 3/12 \ldots . 1/12 \\ 2/12 \ldots 3/12 \ldots . 1/12 \end{pmatrix} \tag{8}$$

In this case we obtain the following structure and return of the portfolio:

$$\bar{d}_{П} = (0{,}083;0{,}388;0{,}529)\ldots \bar{r}_{П} = 14{,}98\%; \sigma^2_{П} = 6{,}4 \cdot 10^{-4}..$$

We consider optimistic option in which the probability of a significant devaluation is twice lower than the probability of moderate devaluation but the probability of moderate devaluation is equal to the sum of probabilities of significant and moderate devaluation:

$$p_{ij} = \begin{pmatrix} 1/12 \ldots 3/12 \ldots 2/12 \\ 1/12 \ldots 3/12 \ldots 2/12 \end{pmatrix} \tag{9}$$

In this case we obtain the following structure and return of the portfolio:

$$\bar{d}_{О} = (0{,}167;0{,}41;0{,}423)\ldots \bar{r}_{О} = 15{,}53\%; \sigma^2_{О} = 7{,}4 \cdot 10^{-4}.$$

The last option is not different from the basic one. In macroeconomic environment and exchange rate instability, the banking system and its clients replace the unstable assets with stable, and this leads to an increase in dollarization of economy in general and the banking system in particular (this quantitative criteria is measured as the share of dollar deposits to the total amount of deposits [5]).

4 Historical Data Model Verification

Model verification can be made on the base of currency exchange rate (UAH/USD) measured for a long period of time and tendencies of the exchange rate of two main world currencies (EUR/USD). For model verification we use period of stable growth of Ukrainian economy from 2002 to 2007 year, which coincides with period exchange rate stability. We calculate the optimal portfolio structure for two periods: after-shock period 2002-2005 and pre-shock period 2006-2007 on the base of NBU data. Average

annual deposit interests for this period is 10%; 5%; 6% and 14%; 9%; 9% (UAH, EUR, USD).

Maximum dispersion magnitude has increased in four times in comparison with previous calculations because of possibility of substantial changes in deposit interests for long period. Optimal portfolio structure has not UAH component in all six possible situation (table 3) for 2002-2005.

Table 3. Optimization of deposit portfolio according to the criterion of profit maximization for 2002-2005 deposit interests: $r_U = 10\%; r_C = 5\%; r_\$ = 6\%$

	θ_1 ($\phi \le 0{,}7$);	θ_2 ($0{,}7 < \phi \le 0{,}9$);	θ_3 ($0{,}9 < \phi \le 1{,}0$)
θ_1^C ($0{,}8 \le \varphi < 1{,}0$)	$d = (0;0;1)$ $S_{t+1} = 1{,}06$	$d = (0;0;1)$ $S_{t+1} = 1{,}06$	$d = (0;0;1)$ $S_{t+1} = 1{,}06$
θ_2^C ($1{,}0 \le \varphi \le 1{,}2$)	$d = (0;0{,}2;0{,}8)$ $S_{t+1} = 1{,}0685$	$d = (0;0{,}2;0{,}8)$ $S_{t+1} = 1{,}0685$	$d = (0;0{,}2;0{,}8)$ $S_{t+1} = 1{,}0685$

Devaluation multiplier UAH/USD probabilities for Tabl.3 ranges calculated from data analisis: $p(\theta_1) = 0{,}077; p(\theta_2) = 0{,}187; p(\theta_3) = 0{,}736$. For EUR/USD devaluation multiplier probabilities: $p(\theta_1^c) = 0{,}449; p(\theta_2^c) = 0{,}551$. Next step probability evaluation of simultaneous occurrence of all 6 possible devaluation states on long time interval:

$$p_{ij} = p_i q_j = \begin{pmatrix} 0{,}035 \ldots 0{,}084 \ldots 0{,}33 \\ 0{,}042 \ldots 0{,}103 \ldots 0{,}406 \end{pmatrix} \tag{10}$$

The expected portfolio structure, for this probability matrix and optimal structure portfolio for each of six situation:

$$\bar{d}_{2004} = (0;0{,}11;0{,}89) \ldots \bar{r}_{2004} = 6{,}47\%, \sigma^2_{2004} = 2{,}5 \cdot 10^{-5}.$$

This result differs from previously obtained for period of crisis. First of all, it concerns the full absence of UAH component, and secondly, much smaller proportion of the contributions in EUR. Both features are explained by ratio of key interests. Difference in interests in UAH was not enough to compensate devaluation risk of national currency, additional interests on USD deposits for EUR provided a small share of EUR deposits.

Optimal portfolio structure for pre-crisis period 2006-2007 differs in increasing share of EUR contribution because interests on USD EUR deposits were equal, UAH share is still equal to zero (Table 4).

The expected portfolio structure for 2006-2007 years:

$\bar{d}_{2006} = (0;0{,}449;0{,}551) \ldots \bar{r}_{2006} = 11{,}4\%, \sigma^2_{2006} = 5{,}2 \cdot 10^{-4}$

Table 4. Optimization of deposit portfolio according to the criterion of profit maximization for 2006-2007 deposit interests: $r_U = 14\%; r_€ = 9\%; r_\$ = 9\%$

	θ_1 ($\phi \le 0{,}7$);	θ_2 ($0{,}7 < \phi \le 0{,}9$);	θ_3 ($0{,}9 < \phi \le 1{,}0$)
θ_1^C ($0{,}8 \le \varphi < 1{,}0$)	$d = (0;0;1)$ $S_{t+1} = 1{,}09$	$d = (0;0;1)$ $S_{t+1} = 1{,}09$	$d = (0;0;1)$ $S_{t+1} = 1{,}09$
θ_2^C ($1{,}0 \le \varphi \le 1{,}2$)	$d = (0;0{,}8;0{,}2)$ $S_{t+1} = 1{,}1336$	$d = (0;0{,}8;0{,}2)$ $S_{t+1} = 1{,}1336$	$d = (0;0{,}8;0{,}2)$ $S_{t+1} = 1{,}1336$

But real structure of bank deposits at that period did not correspond to optimal decision, population prefered UAH deposits because of fixed interests and higher return.

It was thought that the strategy of the fixed exchange rate provided a decrease in the level of dollarization of economy, which is defined as a ratio of foreign currency deposits to all deposits. At this entire interval optimal strategy without risk accounting consists of two key points: borrowing in foreign currency and placing of savings in the national currency. At that time, nobody knew when the period macroeconomic stability would be over, but now it has become clear that the financial crisis was only a trigger for the system that was ready to collapse. UAH savers and currency borrowers who were unable to complete their operations before 2008 crisis had losses. Banking customer behavior on the interval of economic growth can be considered on the basis of the theory of "focusing illusion" [9] when banker clients exaggerate the importance of one factor (fixed course), neglecting the influence of other factors, the effect of which may lead to opposite results.

5 Conclusion

In this research we calculated maximum profitability three components UAH, EUR, USD deposit portfolio structure (targeted function is denominated in US dollars) with risk degree limitations in the economic growth period and periods of macroeconomic instability. The exchange rate instability is regarded as main cause of deposit risks and formalized by the relationship of current currency price to currency price which will be in a year (devaluation multiplier).

Long time devaluation multiplier factor analysis gave possibility to evaluate probabilities of six possible different devaluation (revaluation) situation for pairs UAH/USD and EUR/USD. The optimal solutions were obtained for each of the six possible different situations and for three interest options (two options during economic growth and one during the period of economic turbulence). Expected deposit portfolio was determined in conditions of macroeconomic instability for three possible choices: basic (probabilities of all states are equal), pessimistic (probability of a

significant UAH devaluation is twice higher than the probability of minor devaluation) and optimistic (probability of a significant devaluation is twice less than the probability minor devaluation). For optimistic option the part of UAH deposit must be not more than 17%, in other situation expected UAH part must be not more than 8%.

Optimal portfolio structure in a period of economic grows has not UAH component because of a small difference in the interests of UAH deposits and EUR, USD deposits. But this difference was enough to provide preferred growth UAH denominated deposits. The reasons of this phenomenon is overconfidence of the clients of banking system in UAH stability caused by fixed exchange rate according to NBU strategy.

References

1. Annual report NBU - 2007, online bank.gov.ua/doccatalog/document? id=52855 (2007)
2. Atkinson, A. B., Stiglitz, Joseph E.: Lecciones sobre economía pública. Ministerio de Economía y Hacienda. Instituto de Estudios Fiscales (1988)
3. Bershidsky, L.: Ukraine's Economy Is Worse Than It Looks. online bloombergview.com/articles/2015-03-06/ukraine-s-economy-is-worse-than-it-looks (2015)
4. Bodie, Zvi, Kane, Alex and Marcus, Alan J.: Investments. 7th edition. New York: McGraw Hill/Irwin (2008)
5. Dzyublyuk, O., Vladymyr, O. Foreign capital in the banking system of Ukraine: an impact on the currency market development and banks activity. Visnyk Natsionalnoho banku Ukrainy 5, 26 – 33 (2014)
6. Elton, Edwin J., and Gruber, Martin J.: Modern Portfolio Theory & Investment Analysis. John Wiley&Sons, Inc. (1987)
7. Grushko, V., Ivanenko, T.: Optimization of the structure of the loan portfolio of a commercial bank. Visnyk Natsionalnoho banku Ukrainy, 2, 28 – 32 (2014)
8. Investfunds.ua. Information portal. online investfunds.ua/markets/indicators /usduah- nbu/ (2015)
9. Kahneman, D.; Tversky, A.: On the reality of cognitive illusions. Psychological Review, 103 (3), 582–591 (1996)
10. Kaminsky, A.B.: Modeling of financial risks. Publishing center "Kyiv University", (2006)
11. Cerrato, Mario, Kim, Hyunsok, MacDonald, Ronald: Nominal interest rates and stationarity. Working Papers Business School - Economics, University of Glasgow, online gla.ac.uk/media/ media_150448_en (2010)
12. Monetary and financial statistics, online bank.gov.ua/control/en/publish/article? art_id=67604&cat_id=37801 (2015)
13. Plastun, O., Makarenko, I.: Modeling of the financial markets' behavior during the financial crisis with the use of the fractal market hypothesis Visnyk Natsionalnoho banku Ukrainy, 4, 38–45 (2014)

Dynamic Model of Double Electronic Vickrey Auction

Vitaliy Kobets[1], Valeria Yatsenko[2] and Maksim Poltoratskiy[1]

[1]Kherson State University, 27, 40 rokiv Zhovtnya st., Kherson, 73000 Ukraine
vkobets@kse.org.ua, max1993poltorackii@gmail.com

[2] Taras Shevchenko National University of Kyiv, 90-A, Vasulkivska st., Kiev, 03022 Ukraine
ValeriaYatsenko@rambler.ru

Abstract. The paper deals with different approaches to the definition of e-commerce, including special mechanisms for the distribution of goods and payments, such as auction model. Different formats of auctions that change welfare of their participants are investigated. Software modules were developed for researching the effectivenes of double electronic Vickrey auction. It is defined that in double Vickrey auction incentives for most buyers and sellers are created to reveal their true types. The developed software module of double Vickrey auction showed the highest efficiency in the terms of social welfare among alternative formats and disproved the ability of Vickrey auction to achieve results like market mechanism of perfect competition.

Keywords. e-commerce, online auction, e-auction, Vickrey auction, social welfare.

Key Terms. ElectronicAuction, Software, DoubleAuction, SocialWelfare.

1. Introduction

The current phase of civilization development is characterized by drastic transformations in all spheres of life: from culture and sport to politics and economics. Taking advantage of new methods, changing subject matter of investigations, using neologisms such as digital economy, information economy, info-networks economy, knowledge-based economy, Internet economy, "new" economy, virtual economy, service economy. The variety of modern categories is typical for the modern stage of evolutionary development of international economy, placing special emphasis on the leading role of the triad of determinants of economic growth and development of today, which includes intellectual capital, creative and innovative factor as the basis for developing of knowledge-based economy.

Another feature of the modern epoch of human development is asymmetry of socio-economic development of the international economic system, which is deepening due to globalization. Most scientists think that essential determinants of escalation of global asymmetries lie in ICT-sphere: This leads to more considerable

disproportion in international economy and increases social polarization [1]. 'The Golden Billion' are enjoying their successful development due to unequal relationships with peripheries, as the number of "profitable niches" in global space is highly limited; therefore the way to the civilized "floor" can be easily made due to innovation-information achievements by means of integration the market mechanism into the networked information economy [2].

The technological component of modern economic processes contributes to the development of the networked economy as the synthesis of information and global economies [3].

Works of W. Vickrey, E. Daniel, Gr. Dunkan ,G. Karypis, J. Konstan, P. Cotler, B. Mahadevan, J. Riedl, A. Summer and B Sarwar are devoted to the problems of establishment and development of global and local e-commerce markets in terms of globalization processes. National scientists, namely A. Bereza, A. Berko, V. Vysotska, I. Kozak, F. Levchenko, Y. Lyenshyna, V. Pasichnyk, L. Patramanska, E. Strelchuk, T. Tardaskina do not stand apart of such scientific research. Theoretical and statistical investigations of this category are being conducted by some international organizations such as OECD, UNKTAD, UNISTRAL, WTO and ITU, development projects and strategic programs in regard to e-commerce are being elaborated by World Bank and EBRD.

The paper goal is to ground the impact of e-commerce on the participants' welfare through empirical experiment for electronic auctions, implemented by the means of the relevant transactions via designed software that is economically desirable distribution of goods and payments irrespective of strategic behavior of participants.

The paper has the following structure: the second part is devoted to literature review; the third one determines auction formats; the fourth part constructs the general model of double electronic Vickrey auction for true type and hidden type agents; the fifth part concludes.

2. Related Works

Development of information economy has caused formation of e-society with its integral parts: e-government in politics, e-business in economy, e-education, e-ecology, e-medicine and others (Table 1). E-trading is deemed to be a part of e-commerce which in its turn together with document control and business management makes e-business [4].

In its narrow sense e-commerce is close to e-trading because its main function is online purchase and sale transactions; in the wide sense the definition of e-commerce covers any transaction effected using computer networks [5, 6, 7].

So e-commerce is a complex of business operations carried out using computer networks (Internet, Intranet, Extranet), which are connected with the change of material rights and all processes that support this process including Electronic Data Interchange (EDI), Electronic Funds Transfer (EDF), e-trade, e-cash, e-marketing, e-banking, e-Insurance, e-logistics.

Table 1. Different ways of interpretation the category "e-commerce"

Meaning	Original	Definition
«narrow» (e-commerce=e-trade)	OECD (Ogranization for Economic Cooperation and Development)	An e-commerce transaction is the sale or purchase of goods or services over computer mediated networks (broad definition) or via the Internet (narrow definition) [1].
	R. Doernberg, L Hinnekens, W. Hellerstein, J. Li	E-commerce means the ability to perform transactions involving the exchange of goods or services between two or more parties using electronic tools and techniques.
«wide» (e-comerce = totality of business processes)	A. Sammer, Gr. Dunkan	E-commerce means any form of business process in which the interaction between the actors happens by using the Internet – technology
	UN Experts	E-commerce includes searching for information, contracting, supply of products, goods or services, making payments, sale or purchase of goods or services, whether between businesses, households, individuals, Governments and other public or private organizations, conducted over the Internet. The goods and services are ordered over the Internet, but the payment and the ultimate delivery of the good or service may be conducted on- or offline.
	WTO Specialists	E-commerce is a wide array of commercial activities carried out through the use of computers, including on-line trading of goods and services, electronic funds transfers, on-line trading of financial instruments, electronic data exchanges between companies and electronic data exchanges within a company [2].

Aspects of e-commerce are considered in Table 2 based on [8].

Table 2. Aspects of e-commerce

№	Aspect	Essence
1.	Connections	It is a method of delivery via telephone lines, computer networks, electronic means
2.	Process	It is a technology to automate business operations.
3.	Services	It is a tool to reduce costs, improve quality of goods and services and accelerate delivery.
4.	Time	E-commerce allows to carry out operation online (24 hr. per day).
5.	Space	Open Internet infrastructure makes it a global environment.

According to most experts B2B is the largest segment of e-commerce (Table 3). For example, according to UNCTAD data, B2B is a dominant segment in the American market with twice higher volume of sales compared to those of B2C (559 billion dollars against 252 billion dollars).

[1] Ecommerce Sales Topped $1 Trillion for First Time in 2012. Available www.emarketer.com

[2] E-commerce and Development Key Trends and Issues Available www.wto.org/english/tratop_e/devel_e/wkshop_apr13_e/fredriksson_ecommerce_e.pdf

Table 3. Forms of interaction in e-commerce

Abbreviation	Denomination	Definition
B2B	Business-to-Business	businesses make online transactions with other businesses
B2C	Business-to-Customer	online transactions are made between businesses and individual consumers (social commerce)
B2A	Business-to-Administration	administrative document control
B2G	Business-to-Government	operations between companies and public institutions
e-government	electronic government	e-commerce model in which a government entity buys or provides goods, services, or information to businesses or individual citizens.

It is also confirmed by the structure of the e-commerce market in South Korea based on open sources[3]. (Fig. 1).

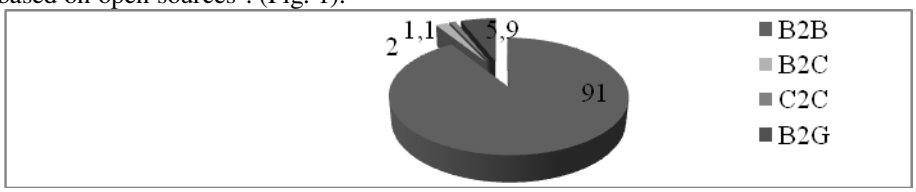


Fig. 1. The structure of e-commerce in South Korea in 2013, %

Internet-shops make an essential part of e-commerce in Ukraine - sector B2C, but B2B segment has great opportunities. For example, International center for electronic trading B2B-center has been successfully functioning for three years in Ukraine, and according to newsb2b.blogspot.com has made it possible to reduce procurement prices for Ukrainian enterprises by 20% on the average as well as procurement labor costs by 70%. The system allows to hold 43 kinds of tender, including more than 172 thousand companies from 110 countries of the world, among them 3500 companies from Ukraine: Group of companies Privat, Ukreksimbank, PUMB, AZOT, Antonov, Ergopak, Rubizhne cardboard mill, Volnogorsk glass, Kolos, Ukrrosmetal, Ukrolia, and international group of companies – JTI, SoftServe, MTC, Allianz. The number of tenders held by Ukrainian segment of the B2B-center system annually increases by 60%. As of today there are two B2B trading sites functioning in Ukraine - b2b-center.ua and b2b-center.uspp.ua, the latter was created by mutual efforts of B2B-center and the Ukrainian Union of Industrialists and Entrepreneurs (UUIE), which allowed it to make online purchases.

The main determinants of insufficient development of e-business and e-commerce are undeveloped technical and technological base. Asymmetric levels of ICT-infrastructure development cause disproportional global development of e-commerce with its traditional key centers in Old and New World – Western Europe and North America, and Asia-Pacific Region (Table 4).

At the same time the growth rates of B2C sales in the developing countries are essentially falling. The highest level is traditionally demonstrated by China (63.8%)[4]. (Fig. 2).

[3] The Statistics Portal. Available www.statista.com

Table 4. Comparative analysis of the development of B2C e-commerce segment all over the world in 2013 [4]

Regions	Sales, bln. dol.	Growth rates, %	Level of coverage, %	Share of sales, %		Deviation %	
North America	419,53	12,5	72	28,3	31,2	2	
Asia Pacific	388,75	23,1	44,6	2,1	2,3		2,9
Western Europe	291,47	14	72,3	26,4	25,4	1	
Central and Eastern Europe	48,56	20,9	41,6	4,1	4	0,1	
Latin America	45,98	22,1	33	34,9	32,9		0,1
Middle East and Africa	27	31	31,3	4,2	4,3		0,2
Total	1,22129	17,1	40,4	-	-	-	-

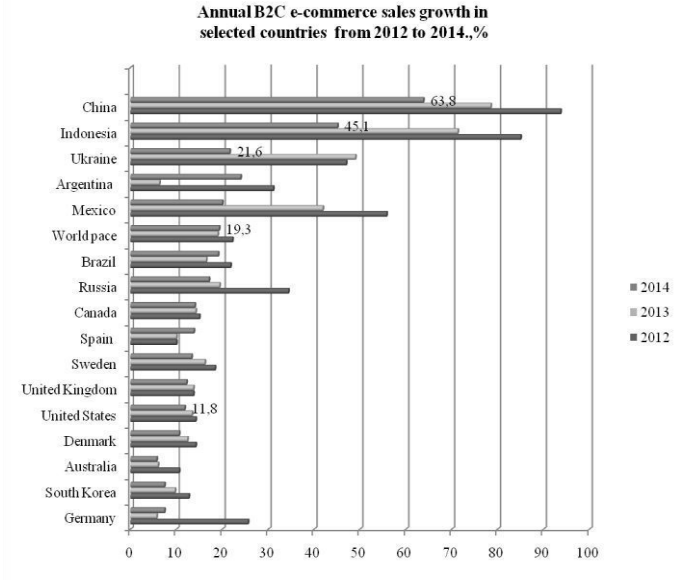


Fig. 2. Growth rates of B2C sales from 2012 to 2014, %

High growth rate at the level of 20.9% is demonstrated by Central and Eastern Europe, but it is lower than in Asia and North America (Fig. 3).

Analysis of commodity composition of e-commerce markets in Ukraine, Russia, Switzerland and the U.S. in 2013 showed disproportional distribution of sales according to segments: e-commerce market in the U.S. is well-balanced and offers a wider range of products than Ukrainian market (Fig. 4). The range of goods in Swiss e-commerce market is not wide but it is well-balanced in contrast to Ukrainian market where 90% of all orders are distributed between two main sectors.

Despite the development of e-commerce business in Ukraine, online-orders do not gain a great popularity with the population, the anticipated level in 2014 was about 3% comparing to 90% in the leading country – the U.S. What makes Ukrainian

[4] Internet business in Ukraine. Available http://ain.ua

market special is that people here mostly use the Internet to learn about the goods, to know about technical specifications of the products, to read other customers' feedbacks, to compare prices and so on, and only a limited number of users place orders, that is why the level of online-shopping and the number of online buyers remain low.

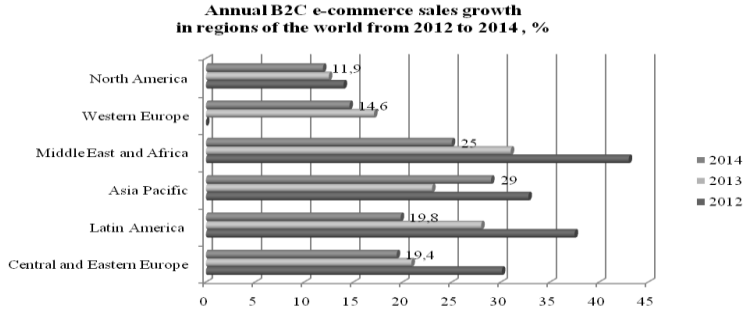


Fig. 3. Dynamics of growth rate of B2C e-commerce sales in regions from 2012 to 2014, % [3]

Fig. 4. The commodity structure of sales in B2C e-commerce segment in selected countries [5].

The results of the research showed the tendency typical for the national markets of all countries – one leading company being in dominant position well ahead of its nearest competitors. Ukrainian e-commerce market is entering the phase of growth because of relatively low volumes of sales of Ukrainian companies (Fig. 5).

In spite of rapid development of e-commerce and e-business in Ukraine there are still some certain difficulties and obstacles that reduce the growth rates of online business as a whole (Table 5).

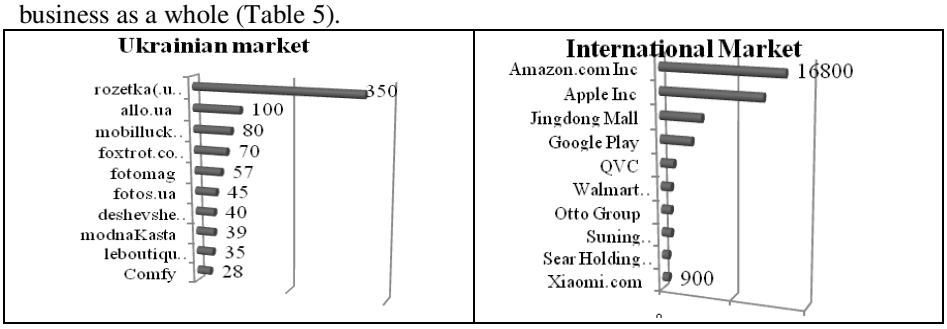


Fig. 5. Comparative analysis of the B2C e-commerce sales volumes in different countries according to investigation in 2013, mln. dollars

Table 5 Problems and prospects of e-business in Ukraine [5]

Factor	Essence
Factors impeding the development of e-business in Ukraine	❖ insufficient development of IT; ❖ limited using of IT; ❖ conservatism and distrust of innovations; ❖ low purchasing power of the population; ❖ lack of specialists; ❖ contractor's mistrust of the banking system; ❖ lack of legal regulation
Prospects for e-business in Ukraine	❖creating jobs for skilled workers; ❖access to Western capital investment; ❖ increasing in tax revenues from the use of electronic payments
Factors accelerating development of e-business in Ukraine	❖ development of electronic payment systems on the Internet; ❖ legislative regulation of the e-commerce, the legal recognition of electronic records and electronic signatures; ❖ protecting commercial information during network transmission

One of the main impacts of e-commerce activity is the formation of certain triad of consequences: product price cutting; speeding up the time and transformation of space (elimination of borders); creation of horizontal links between players and direct contact [4] (Fig. 6).

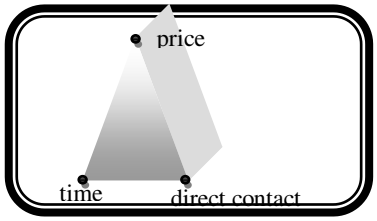


Fig. 6. The triad of e-commerce components

3. Auctions Formats

In most real markets sellers have no perfect information about the market demand, and know only about its statistical distribution. Only buyers know exactly how much product they want to buy at a definite price. Self-regulating market mechanism is not always able to disclose all information about the buyers' solvency and sellers' costs.

The research of decentralized market mechanisms allows us to determine how and why real markets collect and transmit information. Then special mechanisms for the distribution of goods are created, such as auction models.

Auctions can implement the mechanisms of transformation of private information about the value of goods for buyers into common knowledge. In turn, the rules of the auction can stimulate sellers to disclose private information about their cost of goods. Maximum purchasing capacities of the buyer and seller costs are called agents *types*.

[5] Ukraine overview. Available www.ebrd.com/where-we-are/ukraine/overview.html

Designing economic mechanisms for auctions allows building a model of relevant institutions that determine the conditions and means of achieving the goal of the designer [Kobets, 2014]. This model is effective if it allows the planner to create incentives for the disclosure of information held by others to achieve private or public purpose.

To solve these problems auctions mechanisms are designed which motivate the agents to truthfully reveal their private information. Auctions are important for goods that have no natural market, such as bankrupt firms, mobile and radio frequencies. Here accurate information about the number of regular buyers is missing, variance of buyers' values can be very large, and pre-sale valuation and transaction costs are significant.

Operating of Internet-auction is a necessary condition for the development of e-commerce segment and its further growth grounded on [4, 12] (Fig.7).

Effective use of electronic auctions has been confirmed empirically by the most famous giants of global e-commerce such as eBay.com, Sothbys.Amazon.com Yahoo!Auctions and DigiBid.com, which actively use a similar mechanism to promote and sell products and services. Westernbid.com, lotok.com.ua, eTorg.com auctions are gaining their popularity in Ukraine. There are several types of auctions with specific methods of pricing (Table 6):

If a product is sold to the individual who values it most, the auction is efficient. Auction yielding maximum revenue to the seller is optimal [13-14].

Vickrey auction

Agents convey their true type only if it gives them maximum (expected) payoff. Revealing the type means the seller's payoff maximization and efficient allocation of resources (the buyer who values the product most receives it). Vickrey auction (sealed bid second price auction) best of existing auctions formats reveals the types of participants. True strategy is a dominant for Vickrey auction format (as opposed to the first price auction) [15]. The winner receives a payoff as the difference between his own purchase capacity and second price. So when one of the agents has a greater solvency than others, he gets a discount equal to the difference between the first and second largest bids. If Vickrey auction has several winners, then it will select one of them with equal probability.

Then there were 2 extensions of this approach: the *revelation principle*, which showed that direct mechanisms are similar to indirect ones and *implementation theory* that helps to built mechanism so that all its equilibria were optimal ones [16].

Double auction

Theory shows that double auctions, where traders (buyers and sellers) charge their prices can be effective trade institutions, where each agent has private information about his own values of goods.

With the increasing number of traders, the double auction will more effectively generalize personal information so that eventually all information is reflected in equilibrium prices (as argued Wilson). These results are consistent with F.Hayek's argument that markets efficiently summarize relevant private information.

Vickrey auction theory gained wide support from the economists; some elements of the theory have been used in the US in B2G(A) type of e-commerce in organizing the trade licenses to use national radio frequencies. The US State Treasury asked FTC to use this type of auction for revenues maximization.

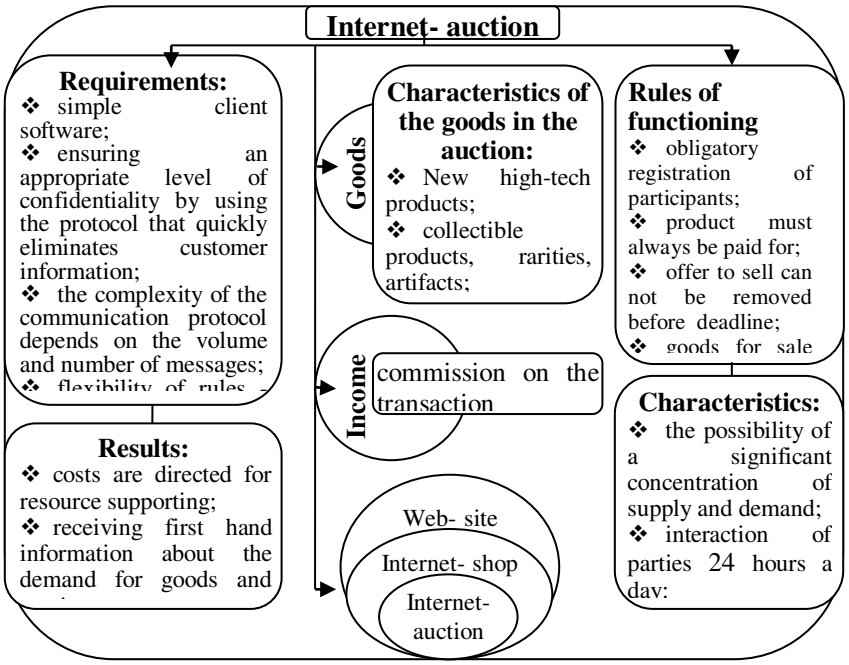


Fig. 7. The mechanisms of pricing in Internet- auctions

Table 6. Diversity of Internet - auctions

Type	Subspecies		Essence
Order	Descending		Next bid is lower than the previous.
	Growing		Next bid must be higher than the previous.
According to the degree of openness	Closed	sealed bid first-price auction	This auction does not disclose the participants' offers. The buyer who offers and pays the maximum price will win.
		sealed bid second price auction	Vickrey auction means the participants don't disclose their proposals. The winner, who has offered the maximum price, pays the second after it.
	Open	English auction	The main characteristic of the auction is that buyers know about competitor's offers. The price starts from a certain minimum level mark. The winner pays the highest price.
		Dutch auction	The main characteristic of the auction is that buyers disclose their bids. The maximum price is fixed and reduced until a buyer agrees to accept it.
		Double auction	The main characteristic of the auction is that buyers and sellers disclose their bids and asks respectively. The seller and the buyer interact the same time - as a result, the equilibrium market price is fixed.

The challenges of the market mechanism require creating rules of interaction for bidders, realized by means of transactions on computer platforms with appropriately developed software and leading to economically desirable distribution of goods and payments deprived of collusion or dishonest behavior of participants.

4. Double Electronic Vickrey Auction Model

To construct the auction model, we introduce the following assumptions. Seller offers one indivisible good to N buyers, who are risk neutral. Buyer i has purchase capacity v_i, $i = 1,...,N$. Evaluation of solvency of buyer i is obtained from the interval [1;100] in accordance with the distribution function $F_i(v_i)$ and distribution density $f_i(v_i)$. Buyers' values of good are mutually independent. Every buyer knows his/her own value and does not know the values of other buyers. However, density distribution functions $f_1,\ldots,f_N$ are common knowledge and are known to both buyers and the seller. Although the seller is uninformed about the exact solvency value of the buyer, he knows the distribution from which each value is received. If the solvency of the buyer who wins the product is v_i, and he pays the price p, his consumer surplus equals $CS_i = v_i - p_i$. The seller's short-run profit will change when the auction format changes.

Sealed bid first-price auction

Buyers make sealed bids b_i that depend on their ability to pay v_i. Buyers' bids are considered as a strategy in the form of functions mapping their solvency in non-negative bid: $b_i \to R_+$. Expected payoff of buyer i will be:

$$CS(r;v) = F^{N-1}(r)\cdot(v-\hat{b}(r)), \tag{1}$$

where r - buyer bid, v - buyer reservation price, $F^{N-1}(r)$ - the probability that the buyer bid on the goods is the highest among all applicants. After first order condition for function maximization (1) and for conditions $F(v) = v$ and $f(v) = 1$ we get size of equilibrium bid for sealed bid first-price auction:

$$\hat{b}(v) = v - \frac{v}{N}. \tag{2}$$

So in this auction format, each buyer conceals his true solvency, relying on a lower bid level than its reservation price.

Double electronic Vickrey auction for true type's agents

Buyers will behave differently in sealed bid first price auction and Vickrey auction. First price auction offers 2 motives for buyer: (i) an incentive to rise his stake to increase his chance of winning; (ii) an incentive to reduce his bid to reduce the price he pays when winning. For Vickrey auction the second motive is not valid, because the winner pays the price which does not depend on his bid. This allows to expect aggressive competition for the good at Vickrey auction. Let B be the second largest bid at the auction, then a winner disclosing his reservation price will win payoff $CS(v) = v - B$.

Suppose that M risk neutral sellers operate in the market. The cost distributions for sellers are obtained from the interval [1;100] in accordance with the known distribution functions. Sellers make sealed asks a_i that depend on their costs c_i. If the seller's cost is c_i, and he gets the price p, his producer surplus (profit) is $PS_i = p_i - c_i$.

Consider our software module for electronic Vickrey auction in Fig. 8.

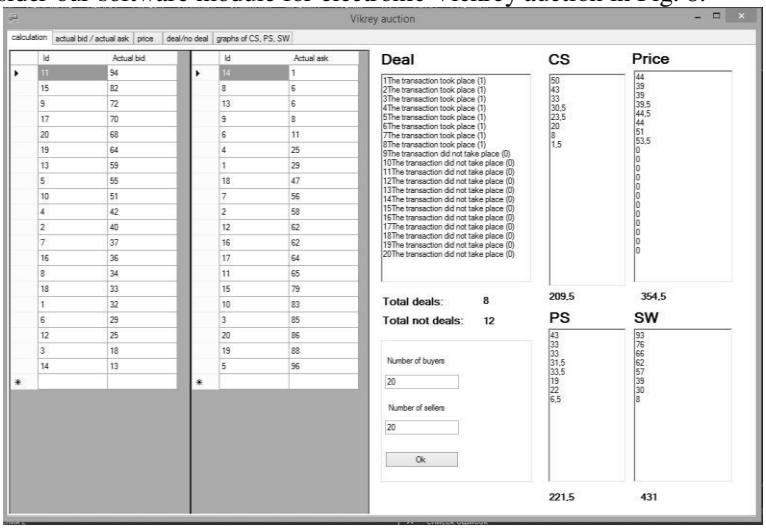


Fig. 8. Double electronic Vickrey auction for true type's agents

In general the number of buyers and sellers may differ $N \neq M$. The buyers' ability to pay is ordered from maximum to minimum and for sellers it is from minimum to maximum.

The agreement between agents (deal = 1) occurs when a price offered by buyer is not below the price set by the seller ($b(v_i) \geq a(c_i)$), otherwise the agents refuse the transaction (deal = 0). The price for each transaction for each pair of buyer and seller is set at the average level:

$$P_i = \frac{v_{i+1} + c_{i+1}}{2}, \tag{3}$$

The auction continues until the highest price offered by a buyer will be lower than the minimum price charged by the seller: $b(v_i) < a(c_i)$ (Fig. 9). After each transaction the benefits of buyers are defined in the form of consumer surplus CS and sellers gains – as producer surplus PS . The sum of consumer and producer surplus forms social welfare SW as efficiency indicator of Vickrey auction format (Fig. 10).

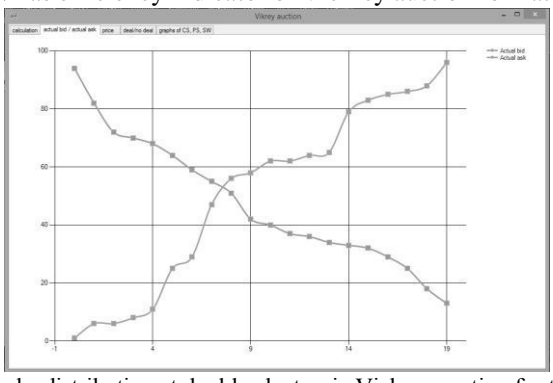


Fig. 9. Bids and asks distribution at double electronic Vickrey auction for true type's agents

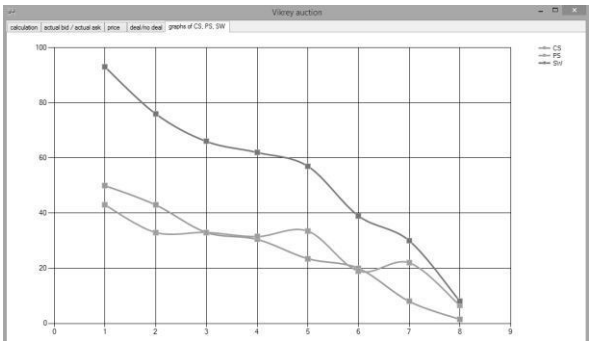


Fig. 10. Dynamics of consumer and producer surplus, social welfare at double electronic Vickrey auction for true type's agents

Fig. 10 shows that functions CS, PS and SW are decreasing in the number of transactions, because during each round of the auction buyers with the highest ability to pay and sellers with lowest cost will benefit. In each round of double Vickrey auction the price of good at first increases and then remains constant, then begins to decrease until it reaches zero (Fig. 11).

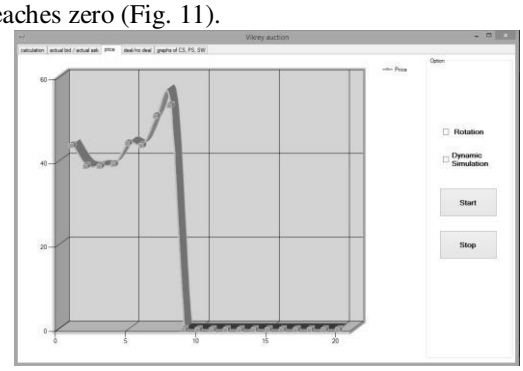


Fig. 11. Price dynamics in double electronic Vickrey auction for true type's agents

Vickrey auction agents' underestimating their ability to pay or overestimating their costs will result in reducing consumer surplus, producer surplus and social welfare. As soon as agents with larger ability to pay and lower cost can not deal in auction, they will discover during few periods that revealing their true type will allow them to maximize their own surplus.

Proposed model of double electronic Vickrey auction for true type's agents is described by the following algorithm by the means of C#:

```
publicvoid Deal(Auction auction)
{intcount_take = 0;
intcount_not_take = 0;
stringpattern_one="The transaction took place (1)";
stringpatter_second="The transaction did not take place
(0)";
for(inti=0;i<auction.Customer_auction.Count;i++)
{if (auction.Customer_auction[i].Bit >=
auction.Seller_auction[i].Ask)
```

```
{count_take += 1;
richTextBox5.Text +=+i+1+pattern_one + "\n";}
else
{count_not_take += 1;
richTextBox5.Text +=+i+1+patter_second + "\n";}}}
publicList<float> Price(Auction auction)
{intprice_did_not = 0;
float average = 0f;
for(inti=0;i<auction.Seller_auction.Count;i++)
{if (auction.Customer_auction[i].Bit >=
auction.Seller_auction[i].Ask)
{average = (float)(auction.Customer_auction[i+1].Bit +
auction.Seller_auction[i+1].Ask) / 2;
averageList.Add(average);
this.richTextBox3.Text += average.ToString()+"\n";}
else if (auction.Customer_auction[i].Bit <=
auction.Seller_auction[i].Ask)
this.richTextBox3.Text += price_did_not.ToString()+"\n";
if (i + 1 == auction.Customer_auction.Count)
break;}
returnaverageList;}
```

Double electronic Vickrey auction for hidden type's agents

During the sale of goods through the auction mechanism a buyer tends to undercharge his own ability, while the sellers tend to overvalue their own costs. So electronic Vickrey auction for true type's agents is less likely than e-auction for hidden type's agents. In theory double Vickrey auction motivates participants to fully disclose their types, because they pay the second largest cost. However, the proposed here new model of double Vickrey auction for hidden type's agents demonstrated that some of agents can hide their true type, despite the existing incentives for disclosure. According to traditional models of double Vickrey auction agent type is disclosed completely.

Consider software module 'Vickrey auction' for double electronic Vickrey auction for hidden type's agents (Fig. 12).

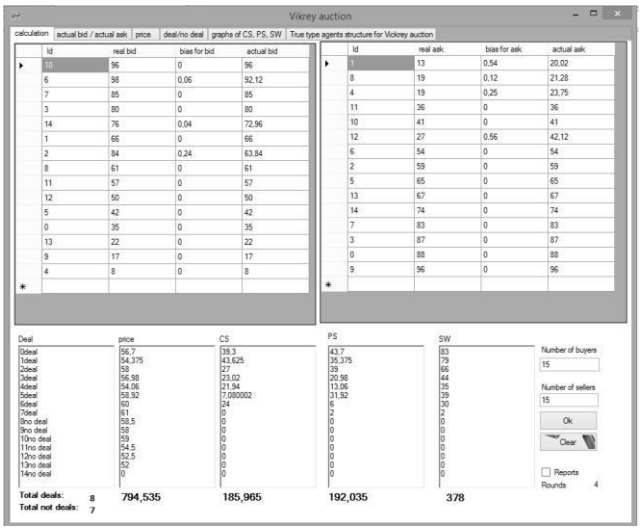


Fig. 12. Double electronic Vickrey auction for hidden type's agents

In this module first we enter *Numbers of buyers* and *Numbers of sellers*. Consider equal numbers of buyers (15) and sellers (15). After entering the data into the appropriate field (Fig. 12), we obtain buyers' real ability to pay *real bid* (as a random number between 1 and 100) and deviation *bias for bid* (as a random number in the interval (0, 1)), which reduces percent of real solvency and gives us actual ability to pay *actual bid*. Therefore the relationship between indicators for buyers looks like: *actual bid = real bid * (1- bias for bid)*. Similarly, we obtain the real costs of seller *real ask* and deviation *bias for ask*, percent of which overstates the actual costs, and reported expenses *actual ask* are received. Thus, the relationship between indicators for sellers is as follows: *actual ask = real ask*(1+bias for ask)*.

After that reported solvency *actual bid* is arranged in descending order, and reported costs *actual ask* is arranged in ascending order. Then pair-wise comparison takes place between the buyer with the highest ability to pay and seller with lowest cost. If $b_i \geq a_i$ then there is an agreement (deal=1) between buyer i and seller i at the price of $P_i = (b_{i+1} + a_{i+1})/2$. For deal i consumer surplus of buyer is $CS_i = b_i - P_i$, producer surplus is $PS_i = P_i - a_i$, social welfare is $SW_i = CS_i + PS_i$. Otherwise, the agreement between the buyer and the seller does not take place (deal=0). Buyers and sellers who do not deal have the incentive to reveal their true types (solvencies or costs).

Proposed model of double electronic Vickrey auction for hidden type's agents is described by the following algorithm by the means of C#:

```
publicList<bool> Deal(Auctionauct, List<bool>DealList)
{intcountdeal = 0;
intcountnodeal = 0;
for (inti = 0; i<auct.Customer_auction.Count;i++ )
{if
(auct.Customer_auction[i].ACtualBid>=auct.Seller_auction[
i].ActualSell)
```

```
{DealList.Add(true);
countdeal++;}
else
{DealList.Add(false);
countnodeal++;}}
returnDealList;}
publicList<float> PS(Auctionauc, List<float>ListPs,
List<float>ListPrice)
{float temp;
floatps;
for (inti = 0; i<auc.Customer_auction.Count;i++ )
{temp = ListPrice[i] - auc.Seller_auction[i].ActualSell;
if(temp>0 &&
auc.Customer_auction[i].ACtualBid>auc.Seller_auction[i].A
ctualSell)
{ps = ListPrice[i] - auc.Seller_auction[i].Ask;
ListPs.Add(ps);}
else
{ps = 0;
ListPs.Add(ps);}}
returnListPs;}
```

But the agreement between buyers and sellers is not completed. Those buyers and sellers who have no deals may revise their bids, that is to reveal their real types. They have an incentive to do so because they haven't got the desired unit of good. After revealing their true type their deviation will be zero: *bias for ask = 0*, *bias for bid = 0*. Further agreements will be revised to reflect the new bids. Then those agents who in the first round were able to buy (sell) goods at their bid and concealed their true type in the second round may lose this opportunity. Then they will get an incentive to disclose their true types. This procedure continues until the final round yields no changes in the redistribution of goods compared to the previous round. It means that the double Vickrey auction for hidden type's agents is completed.

Fig. 12 demonstrates that the buyers' solvency of 2, 6 and 14 remains hidden while the remaining buyers fully reveal their types. Similarly, sellers 1, 4, 8 and 12 did not disclose their true costs, while the rest of the sellers do it. Thus our Vickrey auction model compared with other auctions formats reveals some true types of agents, but this auction format does not motivate all to do as stated in classical Vickrey auction model. In proposed auction model low cost sellers and high solvency buyers can conceal their true types.

For our example 80% of buyers and 73% of sellers reveal their types (i.e. 76% of all traders). 20% of buyers and 27% of sellers conceal their types (i.e. 24% of all traders).

5. Conclusions

To improve e-commerce efficiency there are special mechanisms for distribution of goods and payments such as auctions models that are designed to convert private information about the value of goods for buyers and sellers into common knowledge.

Vickrey auction (sealed bid second price auction) best of the existing auction formats reveals the types of participants. Software modules for dynamic double electronic Vickrey auction were first developed to generalize this auction format. It is determined that in double electronic Vickrey auction incentives are created for *most* buyers and sellers to reveal their true solvencies and costs. But for *some* buyers and sellers these incentives are not enough to disclose their types, which reduces the efficiency of the auction format. The designed program of dynamic double electronic Vickrey auction is closest to perfect competition market and in terms of social welfare ahead of alternative auction formats such as first price auction, English and Dutch auctions, in which the vast majority of agents are hiding and not revealing their types.

References

1. Vdovichen, A. A.: The Causes of Disproportionate Development of the World Economy. Journal of CHTEI. Economics 11 (50), 75--82 (2013)
2. Kovalchuk, T. T.: The Global Information Network Economy: Prospects for Civilization. Urgent Economic Problems 12, 15--23 (2013)
3. Castells, M.: The Rise of the Network Society. Blackwell Publishing Ltd (2e, 2000)
4. Vysotska, V. A.: Features of Planning and Implementation of E-Commerce System. Academic Journals & Conferences of Lviv Polytechnic National University 631, 55--77 (2008)
5. Doernberg, R., Hinnekens, L., Hellerstein, W., Li, J.: Electronic Commerce and Multijurisdictional Taxation. Kluwer Law International (2001)
6. Sarwar, B., Dunkan, A., Karypis, Gr.: Analysis of Recommendation Algorithms for E-commerce. NYH Publishing, 158--167 (2000)
7. UNCTAD E-Commerce and Development Report. Electronic Commerce Branch, United Nations Conference on Trade and Development, Vol. 1 (2001) − Vol. 4 (2004)
8. Pleskach V. L. Zatonska, T. G.: E-commerce. Knowladge, Kyiv, Ukraine, (2007)
9. Bereza, A., Kozak, I., Levchenko, F.: E-commerce. KNEU, Kyiv, Ukraine, (2002)
10. Berko, A., Vysotskaya, V., Pasichnuk, V.: Systems of E-commerce Content. Academic Journals & Conferences of Lviv Polytechnic National University. 612 (2009)
11. Kameneva, M., Gromov, A.: Technology for Virtual Enterprise. Open Systems 4, 155--175 (2000)
12. Pogrebnyak, K. A., Lyenshyna, I. M..: Chameleon Hash in the Group of Points on the Elliptic Curve. Information Processing Systems 3 (93), 129--129 (2011)
13. Yzmalkov, S., Sonin, K., Yudkevych, M.: The Theory of Economic Mechanisms. Problems of Economics 1, 4--26 (2008)
14. Nikolenko, S. I.: The Theory of Economic Mechanisms. Knowledge Laboratory (2009)
15. Nisan, N., Roughgarden, T., Tardos, E., Vazirani, V.: Algorithmic Game Theory. Cambridge University Press (2007)
16. Kobets, V., Poltoratskiy, M: Forming an Evolutionarily Stable Firm Strategy under Cournot Competition Using Social Preferences. In: Ermolayev, V. et al. (eds.) ICT in Education, Research, and Industrial Applications. Revised Extended Papers of ICTERI 2014, CCIS 469, pp. 343-361, Springer Verlag, Berlin Heidelberg (2014) DOI: 10.1007/978-3-319-13206-8_17

Which Data Can Be Useful to Make Decisions on Foreign Exchange Markets?

Karine Mesropyan

Chekhova 41 Rostov-on-Don, Russia 344006
carine@list.ru

Abstract. A communication, settlement of deals, and other services for participants of foreign exchange markets are mostly served by electronic infrastructures. Knowledge of the volume change of aggregated data of deals is useful for all evolved businesses to support their decisions in practice. This paper investigates whether market data of infrastructures, namely CLS, SWIFT, and ETFs, can be used as the volume indicators of some FX segments.

Keywords. Foreign Exchange, Data, Time Series, Review, Flow

Key Terms. DecisionMaking, Management

1 Introduction

The largest and the most influential market for the global and national economies, the foreign exchange market (FX) is opened 24 hours worldwide. According to a regular semi-annual market research an amount of monetary flows traded on the FX in all national currencies (global FX volume) is estimated at 5 trillion US dollar average per day (Bench & Sobrun, 2013). The survey of the Bank for International Settlement (BIS survey) is an important source of the FX knowledge as it aggregates semi-annual surveys from FX committees and provides aggregated statistics of the FX market segments (Fratianni & Pattison, 2001). The monetary amount of trades in every currency is considered in this research as a segment volume indicator of the FX market.

The volume dynamics demonstrates variable numbers which affect exchange rates for national currencies and present volatile level of risk for traders and investors. That is why a kind of the FX volume indicator is a part of system of key performance indicators for assets management among FX market participants. This system helps them in planning and decision making processes such as investment, currency diversification in saving, and development of transnational networks. Market volume is also necessary to be evaluated by national market regulators and central banks in order to have current information on global tendency of their currencies volumes and exchange rates as a consequence.

The concept of volatility is used on financial markets to measure fluctuations of exchange rates by their standard deviation during a taken time interval (Schwartz,

Byrne & Colaninno, 2011; Bubak, Kocenda & Zilkes, 2011). As the volatility of the FX volume is a recognized quantity indicator of the exchange rates dynamics, financial organizations use it in order to estimate their risks on the FX. They usually predict in which extent the exchange rates fluctuate between current level and expiration date.

In order to construct volatility-based market volume indicator, researcher is aimed to investigate area of electronic statistics which is relevant to the FX volume. Several types of time interval can be taken to construct the indicator, namely day, month, and year. The importance of choosing of the time interval for data was emphasized in previous research (Gill, Perera & Sunner, 2012, p.1): "Over recent years technological developments and the digitisation of information and activity have generated a vast array of electronic data, which can potentially be analysed on a daily basis, or even in real time. Some of these data cover very large numbers of individuals and businesses – far more than many traditional surveys used by statistical agencies – and have the potential to be useful for monitoring and measuring aggregate economic conditions."

The BIS survey cannot provide frequent FX statistics, meanwhile market participants become also more interested in monthly volume indicators (Cerutti, Claessens & McGvair, 2012).

Alternative sources of information are investigated in this paper. The FX electronic communication and settlement infrastructures also aggregate statistics on their transactions. Continuously Linked Settlement Bank (CLS) and Society for Worldwide Inter-bank Financial Telecommunication (S.W.I.F.T. or SWIFT) serve financial organizations by secure settlement of their interests on the FX. Exchange traded funds (ETFs) also widely play on the FX market as investment companies which provide efficient and attractive sets of financial instruments in a variety of currencies.

With this research we intend to get a better idea of how the FX market can be measured by using globally aggregated electronic statistics of CLS, SWIFT, and ETFs.

In the first paragraph we state the problem. In the second paragraph we study distinctions and commonality of CLS, SWIFT, and ETFs data regarding to the segments of the FX market. In the third paragraph we investigate relationships between some FX segments volume indicators, namely ETFs in developed currencies and CLS. In the fourth paragraph we discuss methods and our results and in the fifth paragraph we discuss findings and make conclusions.

2 Foreign Exchange Markets in the Last Decades

Developed in 2000s investment opportunities provide a ground for constant enlargement of trades in developing currencies (Bryan, 2008). The latter are reasonably called exotic currencies among the FX practitioners (Tsuyuguchi & Wooldridge, 2008) because they did not find suitable conditions for stable growth worldwide. Thus, from beginning of the post Bretton Woods system in condition of US dollar domination less than 5 % of global trading was made in other local currencies (Pojarliev, 2005). "The relative insignificance of these currencies in

international markets reminds us of the growing disjuncture between countries and "their" currencies. Most Indian- and Chinese-related trade and investment is undertaken in US dollars, with that currency often being used directly without any formal currency conversation (for example, for the purchase of US bonds). Alternatively, for those outside India and China looking for a share of their growth economies, it is possible, using derivatives, to take on exposure to their growth without the need for actual investment in these countries nor for foreign exchange conversion to local currencies." (Bryan, 2008, p. 503).

One could see different environments struggling with implementation of diversification strategies of exchange, saving, and investment in "3 big currencies" and domestic currency. Term of "3 big currencies", namely US dollar, euro, and yen, has become recognized due to trinity's domination in the FX structure (Pojarliev, 2005).

Next, after a crisis of 2008 the global FX market has created a fertile ground for diversification. The post-crisis market conditions have immediately influenced the exchange in a variety of currency pairs, especially in the currencies of developing countries (Bryan, 2008): "…with a declining role for the "big 3" currencies in aggregate, perhaps even the status of any leading national currency being treated as a proxy global anchor is being challenged. Consistent with this trend, it is apparent that foreign exchange is itself being treated increasingly as an asset class (a store of value) as well as a means of exchange, so that investors see intrinsic benefit in holding a wider range of currencies in a diversified asset portfolio."

Nowadays the first candidates to leaders on the FX are Chinese renminbi (yuan) and the Indian rupee which present economies of two members of BRICS (Brasil, Russia, India, China, and South Africa). By World Bank estimation, BRICS contributes a quarter to global domestic product that is more than any other group of developing countries. Although "evolution of the Chinese currency on the FX market remains slow and runs the risk of failing" (Batten & Szilagyi, 2012, p.2), there is an expectation of long-term shift in currency markets. As an evidence of this tendency, New Development Bank (BRICS Development Bank) has been established in 2013 by 5 developing countries as an alternative to International Monetary Fund and World Bank.

Along with currencies diversification a way of presence at the FX is also important for market players. As BIS survey reported, some participants of the FX communicate for trading by using services of brokers but major players replace such supervision by making over-to-counter operations (OTC) themselves. Such market participants are usually members of CLS or SWIFT.

CLS bank serves other banks and financial institutions by mitigating a settlement risk that appears when one party of exchange pays the currency it sold but does not receive the currency it bought (Fisher & Ranaldo, 2011). This kind of risk is called a settlement risk. CLS executes exchange operations (CLS instructions) through provision of its unique payment operation versus payment settlement service. Owing to its service value the CLS is highly appreciated by international financial community (Fisher & Ranaldo, 2011). According to CLS strategy, its large

contribution to the developed markets accompanies by absence on the developing ones.

To serve secure FX transactions, SWIFT plays another role in the industry (Scott & Zachariadis, 2010).

It communicates financial institutions, corporations and their counterparties by SWIFT messages. Their customers are financial institutions, fund managers and brokers, fund managers, settlement members and central settlement systems including CLS members. SWIFT message (MT300) consists of all information about transaction on the foreign exchange such as currency pair, monetary amount, type of trading, and others (SWIFT, 2015).

Nowadays SWIFT possesses a worthy demand in the industry because of its capacity to operate with high value delivered and relatively lower costs in comparison with rivals on both types of markets (developed and developing). This stable trend implies its importance in the industry which provides the SWIFT data potential contribution to the FX volume measurement. Besides, SWIFT does not limit its custodians by a kind of currency to trade. CLS, on the contrary, executes operations in 17 currencies which are mostly developed.

Current tendency on the market is an extremely high growth of ETF segment in both developed and developing economies. Nowadays such indices are traded in a number of currencies on the FX due to its attractiveness for investors.

3 Use of Global Data of CLS, SWIFT and ETFs

CLS data was usually an adequate way of the FX volume estimation. Its monthly market review indicates dynamics of trades on the developed part of the FX. For instance, recently BIS have leveraged CLS information and own data (Bench & Sobrun, 2013). Owing to a mixed approach in monthly numbers measurement, estimated this way FX dynamics was able to explain sources of odd jumps and drops of the market by concrete instruments, which have been described in the BIS survey in a detailed way. Fig. 1 illustrates this first attempt to measure aggregated FX volume for all currencies, including the developed and developing ones owing to the local FX committees' contribution. It makes clear the necessity of the different sources of data combination.

Meanwhile, being outside of mutual work of CLS and BIS, SWIFT could pretend to be considered as a source of data for the FX volume estimation. Its statistics is usually published only in its annual market review where the FX trends are shortly described and illustrated by SWIFT service activity during the year.

Resent research (Cook & Soramaki, 2014) shows that SWIFT data (MT300 message type) is correlated with the FX volume for currency pairs of US dollar and Chinese renminbi (yuan). Authors have found linear relationship between these values (Fig. 2). Absence of similar research of US dollar and yuan from other data sources (CLS, for example) makes impossible to conclude which data is more useful for market analysts by comparing with results of (Cook & Soramaki, 2014) with others.

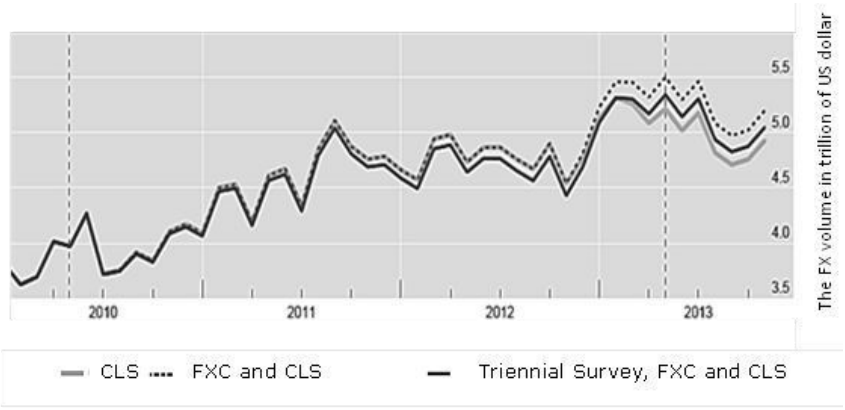


Fig. 1. Monthly CLS data in estimation of global FX volumes in all type of currencies (Bench & Sobrun, 2013)

Alternative source of information comes from investment funds or exchange traded funds which publish their indices volatility. The concept of volatility is used to indicate uncertainty regarding degree of ETFs' volume changes (Schwartz, Byrne & Colaninno, 2011). Although it has been traditionally used for analysis of exchange rates volatility dynamics (Britten-Jones & Neuberger, 2000), we have found examples of its application to measure the range of probable change of traded volume by its dispersion analysis (Melvin & Peiers Melvin, 2003).

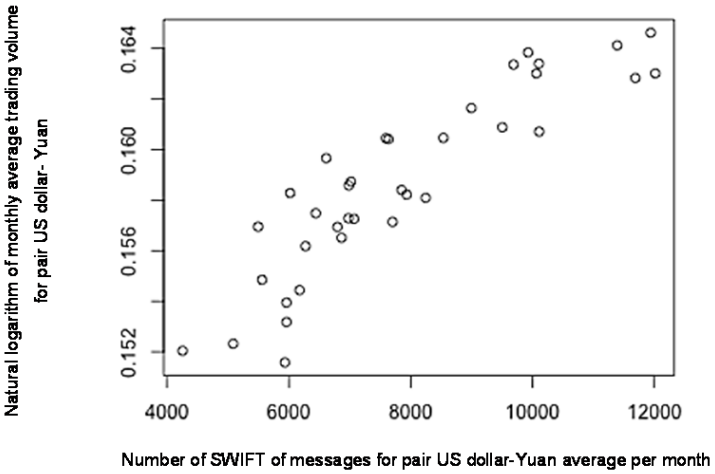


Fig. 2. Monthly relationship between SWIFT data and developed currencies FX segment volume (Cook and Soramaki, 2014, p.27)

As it was reported above nowadays constant growth of the ETF worldwide accompanies by increasing ETF contribution to the financial markets of developing economies. The evolution of financial instruments led to use of ETFs which had

provided implementation of extremely successful trading strategies after the global financial crisis in 2008 (Bryan, 2008, p. 502). As a result, in developed countries, for instance, in the United States ETFs have contributed 40 % of the financial market volume (Guedj & Huang, 2009). As for developing countries, ETFs in currencies have contributed 23 % of the FX market (Fig. 3).

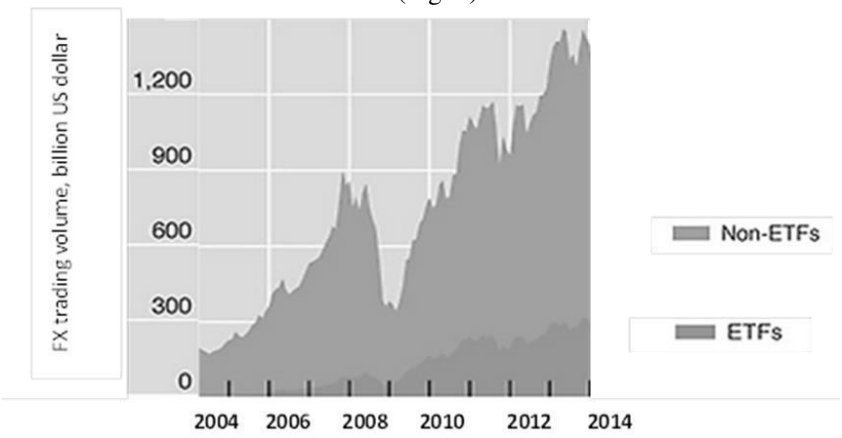


Fig. 3. Yearly ETFs contribution to FX developing currencies segment volume

We have found several studies which are focused on relation between volatility of volume of one of the FX ETF (FXE) and volume of other market segments (Daigler, Hibbert & Pavlova, 2014). Other researchers studied the ETF segment statistics much more widely (Li, Klein, & Zhao, 2012). The common way to construct time series is ARIMA method but the specific for volatility variables is ARCH method (Le & Zurbruegg, 2010). Data about ETFs flows is available on website of Currencyshares' and Powershares' on-line databases which present three largest global currency ETFs indices (US dollar, euro, and yen).

4 Methods and Findings

We have analysed the FX structure by using information from the BIS Survey (2013). According to the survey, traditionally the largest volumes of trading take place in Europe, USA and Japan. Respectively, this trend is presented by the biggest volumes of the FX trading in the main currency pairs, namely Euro versus US dollar (EUR-USD), US dollar versus Yen (USD-JPY), Euro versus Yen (EUR-JPY).

Data have been taken from the website of CLS bank, Currencyshares' and Powershares' on-line databases which present three largest global currency ETFs indices, namely ETF FXE (euro), ETF FXY (yen), and ETF UUP (US dollar).

Data of this research consist of five elements for every month during period from January 2008 till March 2014, namely:

- average number of CLS operations (instructions);
- volatility of volume of exchange-traded fund FXE;

– volatility of volume of exchange-traded fund FXY;
– volatility of volume of exchange-traded fund UUP.

We have studied autocorrelations of these volumes. Table 1 shows that all levels of tested variable of t-statistics are not statistically significant for CLS. According to these results, there are significant autocorrelations between the nearest levels of lags (from month to month) for each time series of ETFs except the time series of CLS.

Table 1. t-statistics of autocorrelation (ACF) and partial autocorrelation functions (PACF)

Lag	ACF				PACF			
№	ETF FXE	ETF FXY	ETF UUP	CLS	ETF FXE	ETF FXY	ETF UUP	CLS
1	6,59*	4,64*	5,74*	1,73	6,48*	4,56*	5,64*	1,62
2	5,37*	4,30*	4,93*	0,29	0,48	2,42*	1,82	-0,31
3	4,34*	3,14*	4,24*	0,95	-0,03	0,13	0,63	0,97
4	3,15*	2,90*	3,10*	0,69	-0,90	0,56	-0,80	0,01
5	2,19*	2,66*	3,16*	-1,13	-0,25	0,61	1,17	-1,32
6	1,28	1,99	3,06*	-1,79	-0,52	-0,35	0,74	-1,03
7	0,29	2,87	2,01*	0,31	-0,82	1,63	-1,33	1,13
8	-0,12	1,68	2,16*	-0,15	0,47	-0,85	0,64	-0,32
9	-0,62	0,86	1,97	-1,05	-0,45	-1,49	0,47	-0,11
10	-1,17	1,02	1,03	-0,59	-0,58	0,76	-1,21	-0,39
11	-1,05	0,92	1,32	-0,09	0,83	0,33	0,43	-0,6
12	-0,69	0,64	1,39	-1,59	0,75	-0,63	0,91	-1,19

Significant levels are signed by (*) on the base of *t*-statistics critical values at the confidence level of 97,5%

Next step should consist of regression models constructing on the base of the time series by using the results of significant lags' autocorrelation. As we have not found out existence of linear relationship between CLS operations from month to month, we could not estimate regression of the CLS and ETFs volumes.

5 Conclusion

A lack of frequently available data can negatively affect strategic decisions of businesses. This research has been motivated by industry's willingness to explain sources of the FX market volume dynamics in developed and developing currencies. In this field we found out several results. Owing to international finance

transformations, nowadays currencies of developing countries become more often used among deals on the FX market than several years ago. This trend had been appeared in the post-crisis period after 2008. The financial organizations had to struggle between two options by making choice on the FX markets. They could adapt to decreasing trends of US dollar domination or they could seize opportunities relating to currencies of developing economies. In 2013 the BIS survey has concluded that unpredictable trends on emerging currencies markets attract more attention of participants to this FX segment volume measurement.

We have also studied what kind of time interval should be taken for the FX volume indicator. We have found that market participants are interested in the FX volume indicators to fill absence of monthly data (Cerutti, Claessens & McGvair, 2012, p.2). Our findings are confirmed by existing statistics source, namely the Triennial Central Bank Survey of Foreign Exchange and Derivatives Market Activity. It collects only long-term overall statistics of volume so the survey cannot respond to need of frequent availability of the FX data without additional market information. That confirms the opinion that "while official statisticians are increasingly using electronic data in the production of economic indicators, this is still very much in its infancy." (Gill, Perera & Sunner, 2012, p.1).

In our review we have concluded that month could be taken as a time interval to make decisions on the FX market by constructing volatility-based volume indicator. Meanwhile, our experimental findings did not provide enough evidence for that.

Next, we have studied the FX data regarding developed and developing currencies. The FX has a number of participants and nowadays only three infrastructures' performances can indicate its overall activity's performance from month to month. Thus, CLS, SWIFT and ETFs data' features analysis has helped to shed a light to data search for the FX volume estimation.

On the one hand, as class of major developed currencies has mostly become an area of CLS business. CLS data can help to measure a volume of trading in currencies of developed countries. CLS does not include trades in currencies of the BRICS countries. South Africa is only one exemption in this group of 5 countries as its currency is considered as a major one and it can be traded by CLS members.

On the other hand, today SWIFT is known as a provider of efficient supply chain for financial organizations in majority of countries including developing ones. SWIFT services are available for all exotic currency pairs on both developed and developing markets. That is why SWIFT membership has become more popular, especially for banks which were not members of CLS.

Potential role of SWIFT information for the FX volume measurement has not been acknowledged yet. Meanwhile, SWIFT has already presented its contribution to economy forecast which was presented by dynamic models for developed (Gill, Perera & Sunner, 2012), developing, and global economies (Bauwens, Gillain & Rombouts, 2011). Thus, SWIFT analytics are more concentrated on current trends of some developing currencies' internationalization such as Chinese Yuan, RMB (Batten & Szilagyi, 2012).

Finally, in our research we have stated a question: 'To which extent does volume of CLS activities indicate the standard deviation of volume during a month for three

major ETF FX segments in developed currencies?' We have calculated the ETF standard deviation on the base of daily volumes in order to aggregate data on the volatility of ETF volume for each month. We have obtained results which have not approved a hypothesis that relation between CLS volumes and ETFs volatility-based estimation of volumes does exist. We were focused on the ETF segment and its three major representatives. These imperfections have affected our research by its inability to extrapolate directly our results for the whole ETF segment of the FX. Next stage of this research could be conducted with more types of ETFs statistics and SWIFT data.

As CLS bank and SWIFT are rapidly evolving competitors in the industry, they consider promotion of the own business intelligence to the FX volume estimation. It makes possible to start research projects in this field. In future research a question can be stated as following: 'To which extent do SWIFT and CLS activities indicate the volume of the major FX segments?' The research objective can be FX volume indicator constructing. Sources of information for the FX size measurement can come from the website of CLS bank, SWIFT, and ETFs (Currencyshares, Powershares, and others) on-line databases.

References

1. Batten, J. A. & Szilagyi, P. G. The Internationalisation of the RMB: New Starts, Jumps and Tipping Points. SWIFT Institute Working Paper, 2012-001. Retrieved from http://ssrn.com/abstract=2325340 (2013)
2. Bauwens, L., Gillain, N., Rombouts, J.V.K. Forecasting GDP Growth Through SWIFT Information. CORE UCL Report. Retrieved from http://www.swift.com./about_swift/shownews?param_dcr=news.data/en/swift_com/2013/PR_index.xml (2011)
3. Bench, M. & Sobrun, J. FX Market Trends Before, Between and Beyond Triennial Surveys. BIS Quarterly Review, December Retrieved from http://www.bis.org/publ/qtrpdf/r_qt1312f.htm (2013)
4. BIS (Bank for International Settlement). Triennial Central Bank Survey of Foreign Exchange and Derivatives Market Activity in 2013. Monetary and Economic Department Working Paper. December. Basel, Switzerland. Retrieved from http://www.bis.org/publ/rpfx13.htm (2013)
5. Box, G. E. P., Jenkins, G. M., Reinsel, G.C. Time Series Analysis, Forecasting and Control. Hoboken: Wiley (2008)
6. Britten-Jones, M. & Neuberger, A. Option Prices, Implied Price Processes and Stochastic Volatility. Journal of Finance, 55, 839 – 866 (2000)
7. Bryan, D. The global Foreign Exchange Market: An Interpretation of the Bank of International Settlements' Survey of Foreign Exchange and Derivative Market activity. Global Society, 22 (4), October, 491 – 505 (2008)
8. Bubak, V., Kocenda, E., Zilkes, F. Volatility Transmission in Emerging European Foreign Exchange Markets. Journal of Banking and Finance, 35, 2829 -2841 (2011)
9. Cerutti, E., Claessens, S., McGvair, P. Systemic Risk in Global Banking: What Available Data Can Tell Us and What More Data are Needed? BIS Working Papers, 12. Retrieved from http://www.bis.org/publ/work376.htm (2012)
10. Cook, S. & Soramaki, K. FX MT300 Correlation Analysis. Mapping Financial Network Analysis (FNA) Report, March (2014)

11. CLS bank. Historical Data from Official Website: http://www.cls-group.com/MarketInsight/Pages/ReportArchive.aspx (2013)
12. Guedj, I. & Huang, J. Are ETFs Replacing Index Mutual Funds? AFA 2009 San-Francisco Meetings Paper. Retrieved from http://ssrn.com/abstract=1108728 (2009)
13. CurrencyShares Euro Trust (Symbol: FXE). Historical Data from Official Website: http://currencyshares.com/products/overview.rails?symbol=FXE
14. CurrencyShares Japanese Yen Trust (Symbol: FXY). Historical Data from Official Website: http://currencyshares.com/products/navs.rails?symbol=FXY
15. Daigler, R.T., Hibbert, A.M., Pavlova, I. Examining the Return-Volatility Relation for Foreign Exchange: Evidence From Euro VIX. The Journal of Futures Markets, 34 (1), 74 – 92 (2014)
16. Fisher, A.M. & Ranaldo, A. Does FOMC News Increase Global FX Trading? Journal of Banking and Finance, 35, 2965 – 2973 (2011)
17. Fratianni, M. & Pattison, J. Review Essay: The Bank of International Settlements: An Assessment of its Role in International Monetary and Financial Policy Coordination. Open Economies Review, 12, 197 – 222 (2001)
18. Gill, T., Perera, D., Sunner, D. Electronic Indicators of Economic Activity. Australian Reserve Bank Bulletin, June (2012)
19. Le, V. & Zurbruegg, R. The Role of Trading Volume in Volatility Forecasting. International Financial Markets, Institutions, and Money, 20, 533 – 555 (2010)
20. Li, M., Klein, D., Zhao, X. Empirical Studies of ETF Intraday Trading. Financial Services Review, 21, 149 – 176 (2012)
21. Melvin, M. & Peiers Melvin, B. The Global Transmission of Volatility in the Foreign Exchange Market. Review of Economics and Statistics, 85, 670 – 679 (2003)
22. Pojarliev, M. Performance of Currency Trading Strategies in Developed and Emerging Markets: Some Striking Differences. Financial Markets and Portfolio Management, 19 (3), 297 – 311 (2005)
23. PowerShares DB US Dollar Index Bullish Fund (Symbol: UUP). Historical Data from Official Website (click on 'historical_navs_uup'): https://www.invesco.com/portal/site/us/financial-professional/etfs/product-detail?productId=uup
24. Qian, B. & Rasheed, K. Foreign Exchange Market Prediction with Multiple Classifiers. Journal of Forecasting, 29, 271 – 284 (2010)
25. Sarno, L. & Taylor M. The Microstructure of the Foreign-Exchange Market: A Selective Survey of the Literature. Princeton Studies in International economics, 89 (2001)
26. Schwartz, R.A., Byrne, J.A., Colaninno, A. Volatility. Risk and Uncertainty in Financial Markets. Springer Science+Business Media (2011)
27. Scott, S. V. & Zachariadis, M. A Historical Analysis of Core Financial Services Infrastructure: Society for Worldwide Interbank Financial Telecommunications (SWIFT). Information Systems and Innovation Group, London School of Economics and Political Science, London, UK. Working paper series, 182 (2010)
28. SWIFT Creating Confidence in a Changing World. Annual Review. Retrieved from http://www.swift.com/assets/swift_com/documents/about_swift/2013_SWIFT_Annual_Review.pdf (2013)
29. SWIFT Harnessing Timely Data for Better FX Decisions. Information Paper. January. Retrieved from http://www.swift.com/assets/swift_com/documents/products_services/Harnessing_Timely_Data_For_Better_FX_Decisions.pdf (2015)
30. Tsuyuguchi, Y. & Wooldridge, P.D. The Evolution of Trading Activity in Asian Foreign Exchange Markets. Emerging Markets Review, 9, 231 – 246 (2008)

Econometric Analysis of Educational Process on the Web-Site

Alexander Weissblut

Kherson State University, Kherson, Ukraine
veits@ksu.ks.ua

Abstract. The paper describes the site "Lesson pulse". It is the tool allowing a teacher to obtain the objective information on the results of a lesson in real-time mode. However, adequate interpretation for the results of such interrogations is impossible while we do not separate true students from the others. Besides, interpretation of the results of interrogations and decision-making grounded on them demand to realize what exactly this specific group means by clearness of explanation, objectivity of marks, etc. For anonymous interrogations it means the necessity of the correlation and regression analysis of the results and an estimation of their statistical significance. So these factors require the use of econometric analysis.

Keywords. Factor, statistical, econometric, analysis, correlation, decision-making.

Key Terms. Research, Management, Model, Knowledge, Management Process, Knowledge Management Methodology, Mathematical Modeling.

1 Introduction

The site "Lesson pulse" is considered in this article. It is the tool allowing a teacher to obtain the objective information on the results of a lesson in real-time mode. However, adequate interpretation for the results of such interrogations is impossible while we do not separate true students, for which educational process is a considerable part of their life, from those who would prefer to keep far away from it [1]. Besides, interpretation of the results of interrogations and grounded on them decision-making demand to realize what exactly this specific group means by clearness of explanation, lesson atmosphere, objectivity of the marks, etc. [2]. For anonymous interrogations it means the necessity of correlation and regression analysis for the results and an estimation of their statistical significance. So these factors require the use of econometric analysis [3].

The site "Lesson pulse" allows a student or a pupil to react to a lesson course at any moment, having answered one or several questions, for example:

1. Is lesson interesting to you?
2. Is the explanation clear to you?

3. Are you tired? Are you satisfied with the rate of the lesson?
4. Do you have some questions to the teacher?
5. Are marks objective?

(Formulations of questions are defined by the teacher) (Fig. 1).

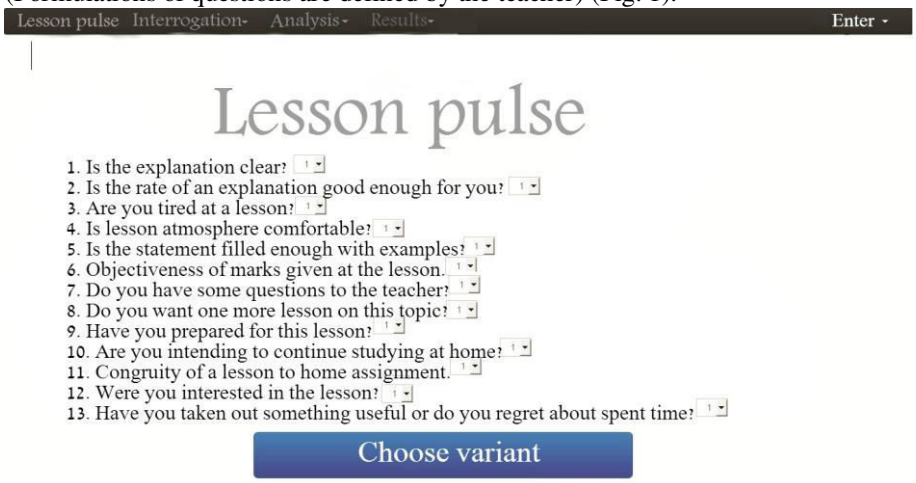


Fig. 1. Lesson pulse

The site displays average marks on responses on the screen. It is the "pulse" of the lesson in real-time mode. At any moment a teacher can ask to answer such or more profound groups of questions (their examples are given below). So, he (she) can measure the “lesson pulse” just at certain moment. Such interrogations do not demand computer auditorium: they can be carried out on a tablet or on a mobile gadget, and then results can be transferred to a site.

1) All groups of questions considered further have been chosen in result of “brainstorming”, where students of fourth year study of the Faculty of physics, mathematics and informatics at the Kherson State University acted as experts. This expert interrogation has been constructed by a technique of “six hats of thinking” by E. Bono [4], which provides the maximal openness and relaxedness of participants. All experts have solidly agreed that this set of questions is full and fair.

2) Then students of speciailties “Physics”, “Mathematics”, “Informatics” and “Software Engineering” of Kherson State University have been interviewed under selected questions. The respondents estimated each question from 0 (at firm “no”) up to 10 (at firm “yes”). He arbitrarily set a name of the folder containing his interrogation (i.e. his key). The volunteer – a participant of interrogation – collects all folders in one main folder and sorts them here (i.e. shuffles). Only after that the main folder is transferred to the teacher: this simple and open procedure guarantee to participants anonymity of interrogation. Alternative and technically simpler variants are answers that are seen on the web-site or could be chosen on a tablet: the variant of choice is defined by the kind of interrogation and the level of trust of an audience to the interviewing teacher.

3) Results of interrogation then are transferred to the site “Lesson pulse”, which is realized in PHP language and uses MySQL database (see [5]). The queries, realizing now on the site, give out results of the econometric analysis of interrogation. They include the plural correlation analysis of factors, the regression analysis and an estimation of the statistical importance of the received results with the use of Student and Fisher criteria ([6]).

The site interface is oriented to the user, generally speaking, knowing nothing about the econometric analysis (Fig. 2)..

Fig. 2. Site interface

2 The Analysis of Interrogations on the Results of a Lesson and Feedback Interrogations

Results of interrogation on a lesson and Feedback interrogation are, of course, absolutely various [7] depending on a lesson, a teacher, an audience, etc (Fig.3).

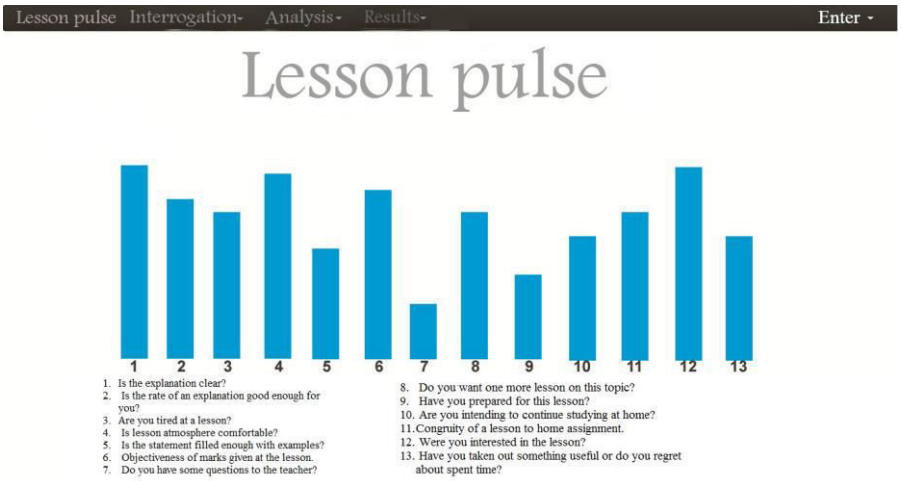


Fig. 3. Results

However, the correlation analysis of factors led to similar outcomes (at 20 % of significance level by criterion of Student). Everywhere below we use the interrogations of the group having typical results on a specialty "Mathematics" (Fig.4).

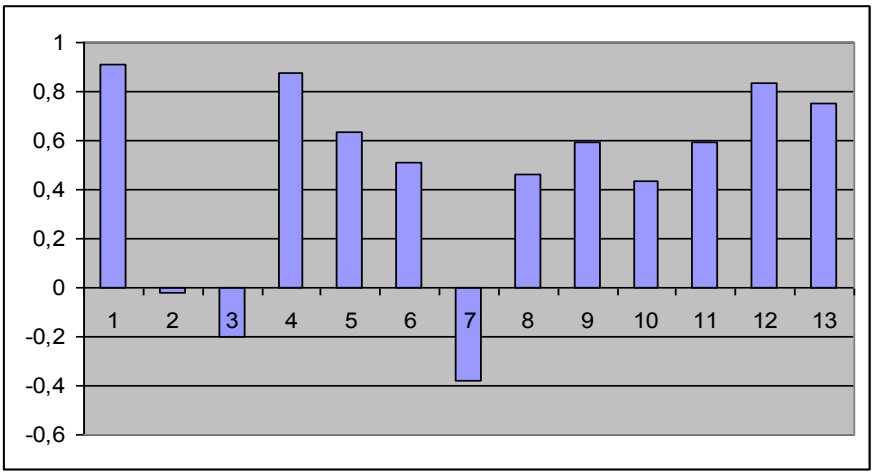


Fig. 4. Questions Distribution

Here is the histogram for distribution of correlation coefficients between answers to a question "Do you like the lesson?" and following factors:

1. Is the explanation clear?
2. Is the rate of an explanation good enough for you?
3. Are you tired at a lesson?
4. Is lesson atmosphere comfortable?
5. Is the statement filled enough with examples?

6. Objectiveness of marks given at the lesson.
7. Do you have some questions to the teacher?
8. Do you want one more lesson on this topic?
9. Have you prepared for this lesson?
10. Are you intending to continue studying at home?
11. Congruity of a lesson to home assignment.
12. Were you interested in the lesson?
13. Have you taken out something useful or do you regret about spent time?

The most significant factors had appeared (in decreasing order) **1** (0,91), **4** (0,87), **12** (0,83), **13** (0,75) **5** (0,63), **9** and **11** (0,59). Objectivity of marks is only further (0,51) and inverse correlation – 0,39 for **7** specifies that a good lesson for the majority is the one after which there are no questions remained to the teacher.

The real importance of examined factors for the lesson estimation is finally established by the regression analysis. At first, we use the most essential factors mentioned above. Then we obtain such linear model:

$Y = 0{,}845454\ x_1 + 0{,}556967\ x_2 + 0{,}32442\ x_3 + 0{,}19571\ x_4 + 0{,}269908\ x_5 + 0{,}24677\ x_6 + 0{,}19877\ x_7$, where the variable x_i corresponds to the factor i ($1 \le I \le 7$). The determination factor for such model is equal to 0, 84572. Using all the factors except insignificant factors **2** and **3**, we obtain the following model:

$Y = 0{,}657012x_1 + 0{,}282476x_4 + 0{,}1349x_5 + 0{,}01807x_6 - 0{,}1097x_7 - 0{,}063159x_8 + 0{,}00809x_9 + 0{,}033186x_{10} + 0{,}126973x_{11} + 0{,}192209x_{12} + 0{,}1266x_{13}$

with the determination factor 0,93647 (Fig.5).

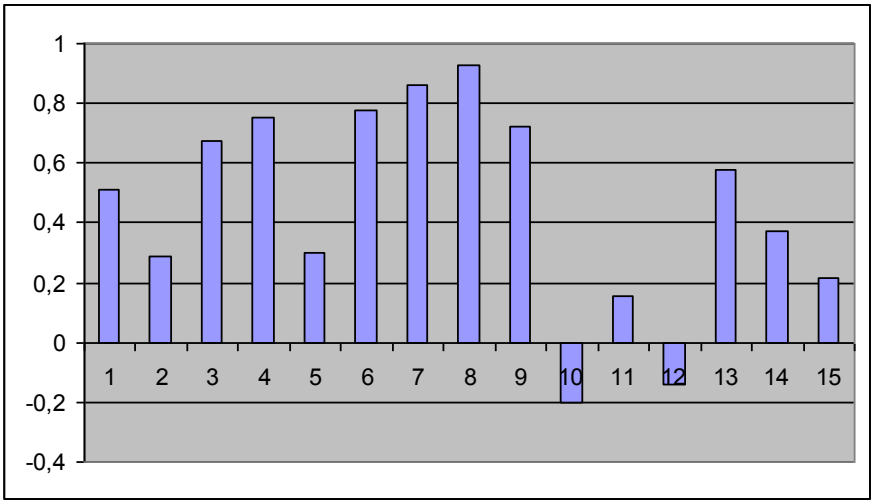


Fig. 5. Questions Distribution

Here is the histogram for distribution of coefficients' correlation between answers to a Feedback question "Do you like your teacher?" and following factors:

1. Do you like the lesson?
2. Student's estimation of the knowledge received at the lesson.
3. Is an explanation clear?

4. Were students' answers clear and adequate?
5. Weather the explanations are filled enough by examples.
6. Using of various approaches during studying.
7. Does the teacher aspire to interest and motivate students?
8. Lessons atmosphere: is it comfortable, is it pleasant to you at the lesson?
9. Availability of the teacher, his inclination to listen the students, to lead a discussion with them.
10. Teacher's competence.
11. Insistence (regular and frequent control of knowledge).
12. Punctuality (comes in time at lessons).
13. Possession of an audience (students are interested in subject and do not make too much noise at the lessons).
14. Objectivity in the teacher's estimation of the student. Are the criteria of estimation in all subgroups identical?
15. Correspondence of the lesson's material to control tasks.

The most significant factors appear (in decreasing order):

8 (0,92), **7** (0,85), **6** (0,775), **4** (0,75), **9** (0,72), **3** (0,675), **13** (0,58).

Only further with factor of correlation 0,51 follows **1** - Do you like the lesson?

Corresponding linear regression model is:

$$Y = 0{,}048604x_1 + 0{,}17976x_3 + 0{,}22221x_4 + 0{,}076545x_6 + 0{,}35703x_8 + 0{,}800305x_9 + 0{,}280308x_{10} + 0{,}23398x_{14} + 0{,}150449x_{15},$$

where the variable x_i corresponds to the factor i ($1 \le i \le 15$). The determination factor for such model is equal to 0,8463.

And major factors of estimations of the teacher and lesson are considerably differing. Further the histogram of differences between factors of correlation for questions "*Do you like your teacher?* and "Do you like the lesson?" is resulted (Fig. 6):

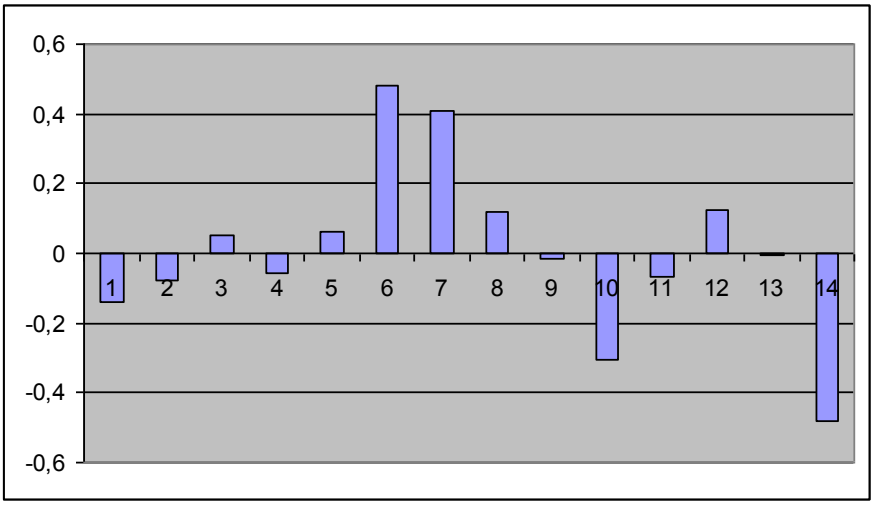


Fig. 6. Questions Distribution

The factors much more essential at an estimation of a teacher, than a lesson are **6** (using of various approaches at training) and **7** (teacher's aspiration to interest and motivate students). On the contrary, at an estimation of a lesson it is much more essential factors **14** (accordance of a lesson's material to control tasks) and **10** – insistence (regular and frequent control of knowledge): probably, according to students, insistence is good at the lesson and it is not so good for the teacher.

Certainly, the correlation matrix contains decomposition on factors also for each of 15 questions. So it is found out that **5** (explanation filled enough by examples) is most closely connected with **15** (accordance of a lesson's material to control tasks); **3** (are you tired at a lesson) with **7** (questions to the teacher); **13** (possession of an audience) with **14** (objectivity in estimation of the student).

It is interesting to compare **12** (is it interesting to you at a lesson) with **13** (have you taken out something useful at a lesson) from interrogation about results of the lesson (Fig. 7).

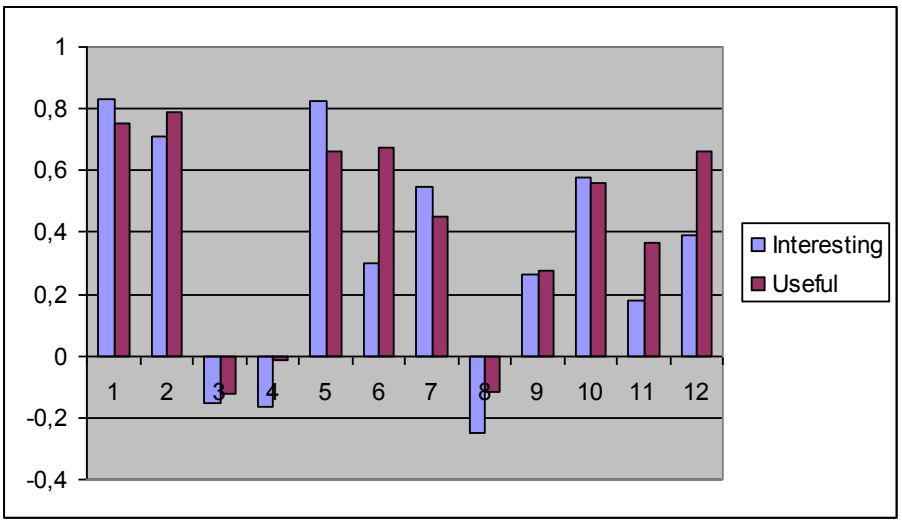


Fig.7. Comparative results

As we see, from the student's point of view, what is interesting and what is useful is not the same. So **4** (lesson atmosphere) correlates with the factor 'interesting', while factor **5** (is the statement filled enough by examples) – with **11** (accordance of a lesson's material to home assignment).

3 The Analysis of Interrogations on the Factors Influencing the Lesson

Unlike interrogations about results of lesson and Feedback results of interrogations about the factors influencing the lesson course [8] are close enough in different

groups. The histogram for distribution of interrogation requisites on the relation to lesson is below (Fig. 8, Fig. 9).

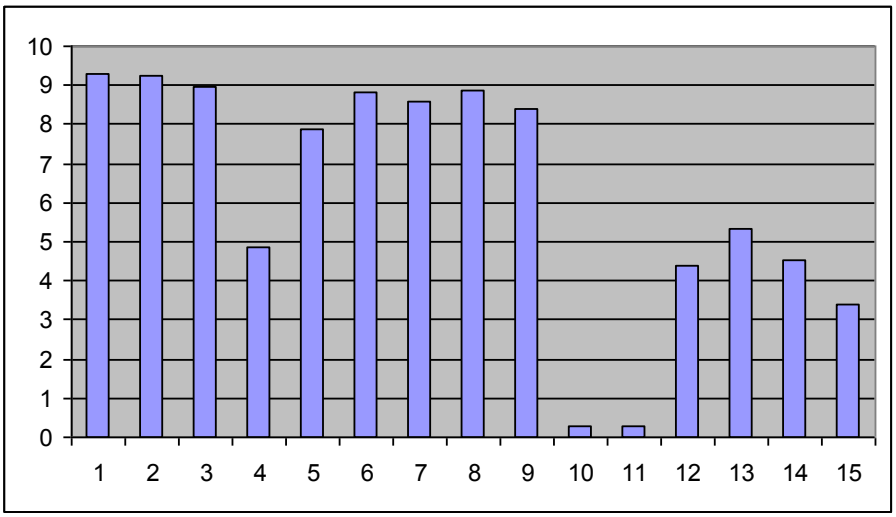


Fig. 8. Questions Distribution

Here:

1. Do you like the lessons? Is the study interesting to you?
2. Do you believe that education is “the road to the future”?
3. Is your speciality interesting to you?
4. Does the training program for your speciality satisfy you?
5. Are you satisfied with your teaching level?
6. Have you chosen university and a speciality on your own?
7. Would you like to change your speciality or enter another university?
8. Do you attend lessons regularly?
9. Are you often prepared with your homework?
10. Did you have any conflicts with teachers?
11. Were you afraid of an elimination from the university?
12. Are you willing to take part in scientific work, in Olympiads on your speciality?
13. How often do your classmates address to you for the help?
14. Do you wish to enter postgraduate study after you studying ends?
15. How much time do you spend for preparation for lessons (hours per day)?

Similar results of interrogation on external factors are the further:

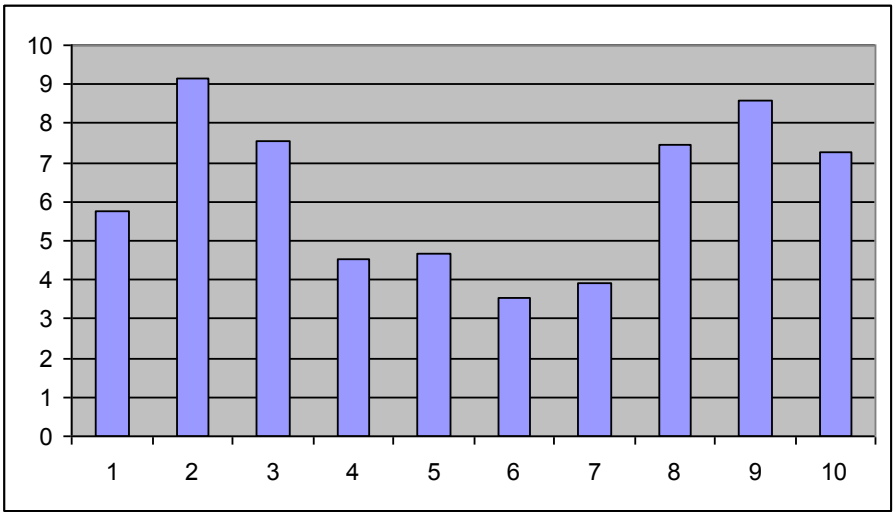


Fig. 9. Questions Distribution

Here:

1. Close interaction with teachers.
2. Accessibility of the Internet at university.
3. Preparedness of an auditorium for a lesson (working projectors, computers, the software; comfort of an auditorium).
4. Presence of enough points for the centralized feeding.
5. Accessibility of contacts with the future employers.
6. Accessibility of summer recreation.
7. Participation in scientific work.
8. Teaching level at the university.

In a correlation matrix under all these factors there are only few factors which correlations are close to 1. These are factors:

1. Do you attend lessons regularly? with factors
   1) are you often prepared with your homework (0,87)
   2) teaching level at the university (0,84)
   3) participation in scientific work (0,63)
   4) have you prepared for this lesson (0,59)
   5) accessibility of summer recreation (– 0,5).

2 Are you often prepared with your homework with factors
   1) do you attend lessons regularly (0,87)
   2) teaching level at the university (0,815)
   3) have you prepared for this lesson (0,66)
   4) participation in scientific work (0,56)
   5) accessibility of summer recreation (– 0,52).

3 Teaching level at the university with factors
   1) do you attend lessons regularly (0,843)
   2) do you regularly prepare homework (0,815)

3) have you chosen university and a speciality on your own (0,65)
4) have you prepared for this lesson (0,59)
5) participation in scientific work (0,56)
6) accessibility of summer recreation (– 0,55).

Besides them correlation factors above 0,7 appear still only twice: between factors *Did you have any conflicts with teachers* and *Were you afraid of an elimination from the university* (0,85); and between factors *participation in scientific wor*k *and Are you satisfied with your teaching level* (0,74). Occurrence in such line the *factor teaching level at the university* is, probably, the best compliment for Faculty of Physics, Mathematics and Informatics of the Kherson State University for all its history. Our main task is to use the mental orientation, fixed thus in the correlation analysis of factors, for separating true students, for which educational process is a considerable part of their life, from those, who would prefer to keep far away from it. Using already cited data and the following table 1:

Table 1 Data

Factor	Average value	Root-mean-square deviations
Teaching level at the university	7,2	2,17
Regularly attendance of lessons	8,85	2,3
Regularly prepare homework	8,4	2,6

we choose as a differentiating sign between groups the factor *regularly of homework preparedness*. In this case mutual correlations of defining sign are closer to 1; and the dispersion is more, that testifies about more variability of respondents under this factor. Besides, among others selected it corresponds more to such sign in common sense.

4 Results of Interrogations about Lesson and Feedback on Subgroups

To the selected differentiating sign among 20 respondents of group the 12 participants is allocated, who for a question *Are you often prepared with your homework* have answered with 10 or 9 points. The additional subgroup consists of 8 respondents. Do such subgroups correspond to required division into true students and the others? Below there is the histogram for average results of interrogation on the lesson on the allocated subgroups (Fig. 10, Fig. 11).

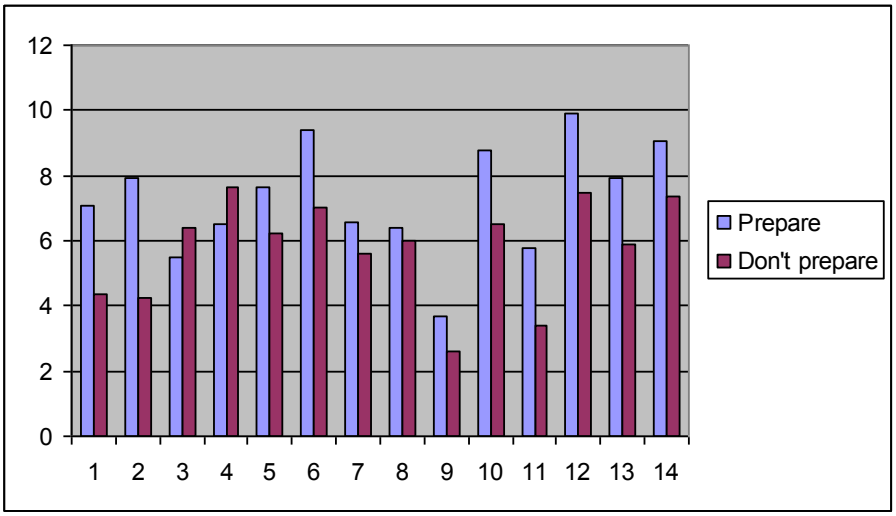


Fig.10. Comparative results

So, the factors considerably different in subgroups (in decreasing order of modules of differences between average values in subgroups) are:

2 Is an explanation clear? (7,92 – 4,25 = 3,67)
1 Do you like the lesson? (7,1 – 4,38 = 2,72)
12 Accordance of a lesson to home assignment. (9,91 – 7,5 = 2,41)
6 Is the statement filled enough by examples? (9,41 – 7 = 2,41)
10 Have you prepared for this lesson? (8,75 – 6,5 = 2,25)
13 Is it interesting to you at a lesson? (7,92 – 5,87 = 2,05)
14 Have you taken out something useful at a lesson? (9,1 – 7,4 = 1,7)
5 Lesson atmosphere (7,66 – 1,25 = 1,41)
9 Do you want one more lesson on this topic? (3,66 – 2,65 = 1,01)

The averages of additional group are more only twice, there are:

4 Are you tired at a lesson? (6,5 – 7,62 = – 1,12)
3 Is the rate of an explanation good enough to you? (5,5 – 6,37 = – 0,87)

Last result seems strange at first sight, but it is steady for all groups and it is easy to explain this phenomenon psychologically: the less the student is adjusted for the study, the more he would like to speed up lesson's time.

Further there are similar results for Feedback interrogation.

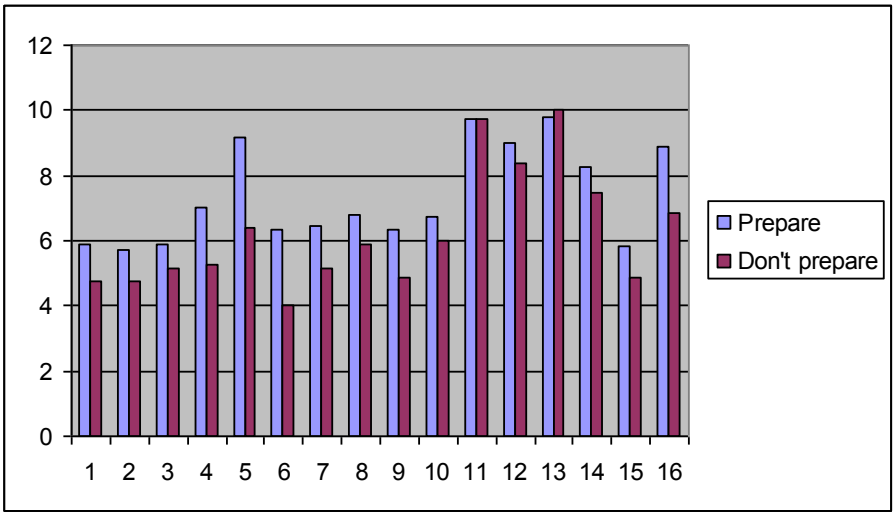


Fig.11. Comparative results

Here are the factors considerably different in subgroups:

5 The explanation is filled enough by examples (9,17 – 6,37 = 2,8)
6 Using of various approaches at studying (6,36 – 4 = 2,36)
16 Accordance of a lesson to control tasks (8,9 – 6,87 = 2,03)
4 Are the answers clear enough? (7 – 5,25 = 1,75)
9 Lesson atmosphere (6,36 – 4,85 = 1,51)

The obtained data corresponds to a hypothesis about required division into groups, anyway they don't contradict it.

5 The Latent Division in Group

The site "Lesson pulse" offers also group division into classes with a given value of mutual correlation: between two respondents from one class it is possible to find a chain of respondents of this class in such a way, that the correlations of answers between consecutive respondents of this chain is not less than the given value. Such division into subgroups allows finding out distinctions in the group, which are not appreciable directly.

At mental interrogation about factors of influence on lesson and the set minimum level of mutual correlation 0,6 in test group 421 splitting into 3 classes has turned out: from 4, from 5 and from basic subgroup of 11 respondents. Let's compare averages of the basic class to averages of the first and the second subgroups under those factors in which appreciable differences have come to light (Fig. 12).

1) Is the program of training for your speciality satisfying you?
2) Would you like to change the speciality or enter another university?

3) Are you willing to take part in scientific work, in Olympiads on your speciality?
4) Do you wish to enter postgraduate study after training end?
5) Participation in scientific work.
6) Preparedness of an auditorium for a lesson.
7) Accessibility of summer improvement.
8) Accessibility of contacts with the future employers.

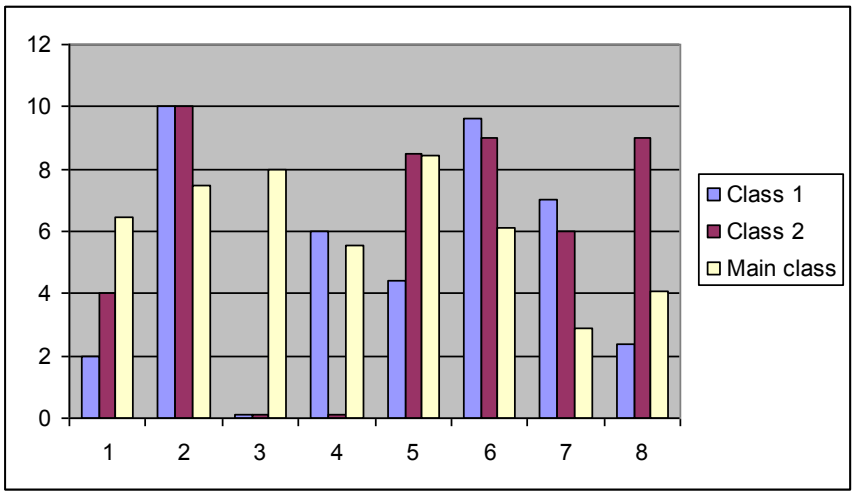


Fig.12. Comparative results

Respondents from classes 1 and 2 much less than the basic group are satisfied by the program of training of the speciality (point 1). They would like to change the speciality or to receive additional higher education much more than the basic group (point 2). Their difference clearly comes to light in point 3: unlike the basic group they do not wish to take part in scientific work or in the Olympiads on the speciality at all. So, apparently, the speciality has lost now its appeal for them. Respondents from class 2 are not interested in the postgraduate study (point 4), however, they are not against taking part in scientific work (point 5). The main thing, they have the most interest in contacts to employers (point 8). Apparently, it is search for their employment out of the speciality. Respondents from class 1 are focused differently: they have a little interest in scientific work and employers (points 5 and 8), but they wish to enter postgraduate study (point 4).

References

1. Research Spotlight on Academic Ability Grouping (NEA Reviews), http://www.nea.org/tools
2. Jennifer J. Kaplan, John G. Gabrosek, Phyllis Curtiss, and Chris Malone Investigating Student Understanding of Histograms, Journal of Statistics Education 22(2) (2014)

3. Greene, William H. Econometric Analysis, Prentice Hall. (2012)
4. De Bono E. Six Thinking Hats. Penguins Books. (1997)
5. PHP Book, http: // www.phpreferencebook.com/
6. Hansen B. E. Econometrics (2012), http: //www.ssc.wisc.edu
7. Pam Boger Building the Numeracy Skills of Undergraduate and Elementary School Students, Journal of Statistics Education 13(3) (2005)
8. Factors Affecting Learning, http://www.gdrc.org/info-design

The Multidimensional Data Model of Integrated Accounting Needed for Compiling Management Reports Based on Calculation EBITDA Indicator

Yatsenko Viktoria

Kherson National Technical University, 24, Beruslavske st., Kherson, 73008 Ukraine
Viktorijajacenko@rambler.ru

Abstract. Organization and method of assembly management report using a definition of EBITDA indicator (Earnings before interest and tax, amortization and depreciation) are considered on the practical example of Kherson river port's activity. The process of constructing a multidimensional model, that is necessary for determining EBITDA of integrated accounting using the program "1C: Accounting for Ukraine", and implementation of the model using PivotTable in MS Excel are represented. The range of possibilities to implement the process named "Data Mining" of the models is demonstrated. The management report, formed on the basis of multidimensional data model is used to determine the profitability of the business units, business processes and enterprises considering the organizational architecture of the entity.

Keywords: Multidimensional data model, EBITDA, Pivot table, Management Reporting.

Key Terms: KnowledgeManagementMethodology, Management, Model, ModelBasedSoftwareDevelopmentMethodology.

1 Introduction

The process of the evolutionary development of accounting and reporting in Ukraine has a long history of changes and qualitative transformations, first of all, resulting from the wish to timely provide the various groups of interested users with the reliable data. The accounting and financial reporting is considered by lots of people as a formally obligatory phenomenon approved and regulated by the state legislation. Actually, it is a fundamental basis of the de facto existing accounting and analytical system able to perform the primary functions of a business management.

The current realia of a business management require expending the boundaries of the existing accounting and reporting systems be means of including the tasks of planning, control and performance measurement regarding the activities, business processes, business units, the company as a whole, and elaborating the strategy of operation and development. Additional "non-standard" for the accounting requests

from the information users and different vision of the functional tasks' development essentially enforces the formation of various types of accounting (financial, fiscal, management, strategic, etc.) and the methods of data interpretation in order to define the financial indicators such as (EBIT, EBITDA, ROA, TIER etc). One of the priority trends of the accounting development is creating the accounting and analytical system of a company that can provide all the necessary information to every level of management on a real-time basis. In this paper we present a variant of an integrated accounting data model on a company incomes and expenditures allowing you to create the management report items based on the indicator computation EBITDA, and which is in practice used at Kherson river port.

2 The System of Integrated Accounting

New approaches to shaping the views of the category "accounting" are based on the theory of the system: any system can be represented as a set of the inter-related and linked elements forming a certain unity and value. In addition, it is necessary to emphasize the impact of the system theory on understanding the accounting system as a multidimensional and complex informational space. Determination of the core system features is an important factor allowing seeing the accounting elements in a single accounting system. These features include: the ability to assess data to solve problems in the same monetary units; matching the economic resource cycle model, their origin, and business processes represented in the general Chart of accounts; actuality and retrospectiveness of the data obtained within the framework of accounting; legal (documentary) proofs of business transactions. The system approach to formation of the indicators to draw the various forms of reporting (financial, statistical, fiscal, management) makes it possible to assert of the establishment and operation of an integrated accounting system.

In recent years, the problem of integration of the accounting information has become particularly relevant for the scientists. A lot of them raise the issue of the necessity to get the information that allows separating the costs not only for the reproducing process as a whole, but also for all types of the core and service processes which is important to identify the most costly processes, develop measures to reduce the costs for their implementation [1].

The integrated accounting is the main element of the accounting and analytical business management and the basis for the accounting system functioning that allows you to transform information in order to draw various forms of reporting and identifying the indicators characterizing the degree of the approved plans implementation.

Analysis of the possibilities of the special-purpose programs of various decision support systems (DSS) available on the Ukrainian market confirms that the software products meet the requirements put forward by the modern company executives and enable to simulate any business processes with due consideration of the external and internal factors, and can automatically calculate the economically sound company's performance indicators. The main criteria for choosing the software for the Ukrainian companies is the minimum price, usability, compatibility with accounting programs usually on the 1C platform, and preferably not involving any IT experts.

3 The Multidimensional Data Model

Let us give consideration to the real-life experience of the Kherson river port on solving the tasks mentioned.

Evaluation of the performance of the Kherson river port is based on the EBITDA indicator, which, according to the foreign authors, is the key to determine the profitability, and is used all around the world [2].

The indicator EBITDA (Earnings before interest and tax, amortization and depreciation) means Earnings before interest, taxes, depreciation and amortization [3]. There are several algorithms for calculating EBITDA. The company in question uses the following order to calculate the analytic indicator, as adapted to the realities of its economic activity:

$$EBITDA = NP + ITE - SIT + IE - EI + AA - DA, \quad (1)$$

where NP - net profit, ITE - income tax expense, SIT - satisfied income tax, IE - interest expenses, EI - earned interest, AA - amortization of assets, DA - depreciation of assets.

Kherson river port maintains the financial accounting and prepares financial statements pursuant to the national Regulations (standards) of accounting (NP(S)A) and the International Financial Reporting Standards (IFRS) in parallel, which meets the requirements of the Law of Ukraine on Accounting and Financial Reporting [4]. It is clear that the definition of EBITDA is not possible on the basis of the financial accounting data without further transformation.

The process of accounting and financial reporting at the company in question is automated using "1C: Accounting for Ukraine". Necessary details of the accounting data in "1C Accounting for Ukraine" as the raw data to determine the resulting indicator EBITDA is achieved by constructing a hierarchy of the analytical accounting levels through the structured directories for storing objects that can be hierarchically classified according to selected features. Important for the determination of EBITDA and preparation of management reports is the organizational structure, under which one should understand a complex of the typical elements of accounting in general and some of its parts in particular. Given the category features and integrated accounting, to build a multidimensional data model of EBITDA determination, a basic scheme of the integrated accounting of income and expenses is used at the company in question on the "asterisk" principle (table 1).

Construction of the model takes into account the complex organizational architecture of the company as well as the details of its activity. The point at issue is that the business units of the company are strongly interrelated and also perform the maintenance functions of the company in general, therefore, it is important to separate the data relating to the internal business volume to prevent any result misrepresentation.

Table 1. The multidimensional organization model of the integrated accounting of the Kherson river port.

Characteristic	Period			Subject	Object	Area	Dimension
Dimension of characteristics			Month	Business Unit (BU activity)	Impact on the result of activity (+/-)	Type of activity	Data unit
Meaning of characteristics	Year	I half-year	I quarter	January	Port	loading and unloading operations	Internal turnover	Income	Operating Financial Another	Result of dimension
				February		Elevator				
				March						
			II quarter	April		Complex fleet service				
				May		Mechanization				
				June		Cargo and passenger services				
		II half-year	III quarter	July		Another				
				August						
				September		Sand		Costs		
			IV quarter	October						
				November		Non-core assets				
				December						

4 Implementation of the Model in Pivot Table MS Excel

The process of drawing a management report based on EBITDA for the Kherson river port is realized in the pivot Excel tables, "... one of the most convenient means applying the OLAP technology, the main purpose of which is to process information for analyzing and decision making. The advantage of OLAP is to create queries using flexible ad hoc approaches without involvement of the IT experts. The pivot tables

provide using of the multidimensional classifications, detail and integration of the data, identifying trends, patterns, forecasting, analysis, thus representing a weighty tool for operation of the accounting and analytical system of a multi-segment company in the real-time mode" [5].

Formation of items of the management statements based on the multidimensional data model of the integrated accounting and the algorithm for determining the EBITDA indicator in the pivot Excel tables are shown in Table 2,3.

Table 2. The management report items in the Pivot Tables in MS Excel of the Kherson river port.

Status	The management report items	Abbreviation
(=)	Total revenue	TR
(-)	Logistics costs	LC
(-)	Special engineering	SE
(=)	Present revenue	PR
(=)	Variable costs	VC
(=)	Marginal revenue	MR
(=)	Marginal revenue, %	MR, %
(-)	Material costs	MC
(-)	Energy	E
(-)	Insurance	I
(=)	Services of external organizations	SEO
(-)	Staff costs	SC
(-)	Depreciation	D
(-)	Change of residues unfinished goods and finished goods, corrections of balance residues	CBR
(-)	Operating taxes	OT
(=)	Fixed costs	FC
(=)	Total profit	TP
(+)	Other income	OI
(-)	Other costs	OC
(=)	Profit before taxes 1	TP1
(+)	Financial income	FI
(-)	Financial costs	FC
(=)	Profit before taxes 2	TP2
(-)	Income tax	IT
(=)	Net profit	NP
(=)	Net profit, %	NP,%

Table 3. The calculating algorithm of performance indicators of the management report in the Pivot Tables in MS Excel of the Kherson river port.

Indicators	Abbreviation	Algorithm for calculating
Total revenue	TR	*SUM* (TR_ cargo fleet: TR_ transportation fleet ports; TR_ cargo handling; TR_ comprehensive fleet maintenance; TR_ rental income; TR_ industrial activities; TR_ non-core activity; TR_ other income)
Present revenue	PR	*SUM* (TR; LC; SE)
Variable costs	VC	*SUM* (v_ fuel; v_ material costs; v_ port charges; v_ energy; v_ taxes)
Marginal revenue	MR	*SUM* (PR; VC)
Marginal revenue, %	MR, %	*IF* (TR=0;0;MR/PR)
Services of external organizations	SEO	*SUM* (f_ assignment; f_ repair; f_ rent; f_ connection; f_ other costs)
Fixed costs	FC	*SUM* (f_ material costs; f_ energy, f_ insurance; f_ services of external organizations; f_ staff costs; f_ depreciation; f_ corrections of balance residues; f_ operating taxes)
Total profit	TP	*SUM* (MR; FC)
Profit before taxes 1	TP1	*SUM* (TP;OI;OC)
Profit before taxes 2	TP2	*SUM* (TP1;FI;FC)
Net profit	NP	*SUM* (TP2; IT)
Net profit, %	NP,%	*IF* (PR =0;0; NP / PR)
EBITDA		*SUM* (NP; IT; FC; FI ; D*)*

5 Capabilities of the Model

A management report implemented in the pivot tables represents the data as to several informational slices - forming a subset of a multidimensional amount of data corresponding to one or more elements of measurement. For example, selecting a subset of values of the company's fixed costs over certain time, as a structural unit in general, and those of a business unit in particular, highlighting the internal business volume (fig. 1).

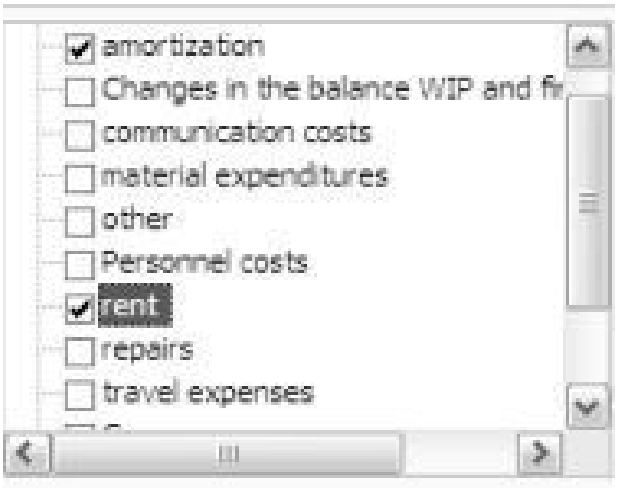
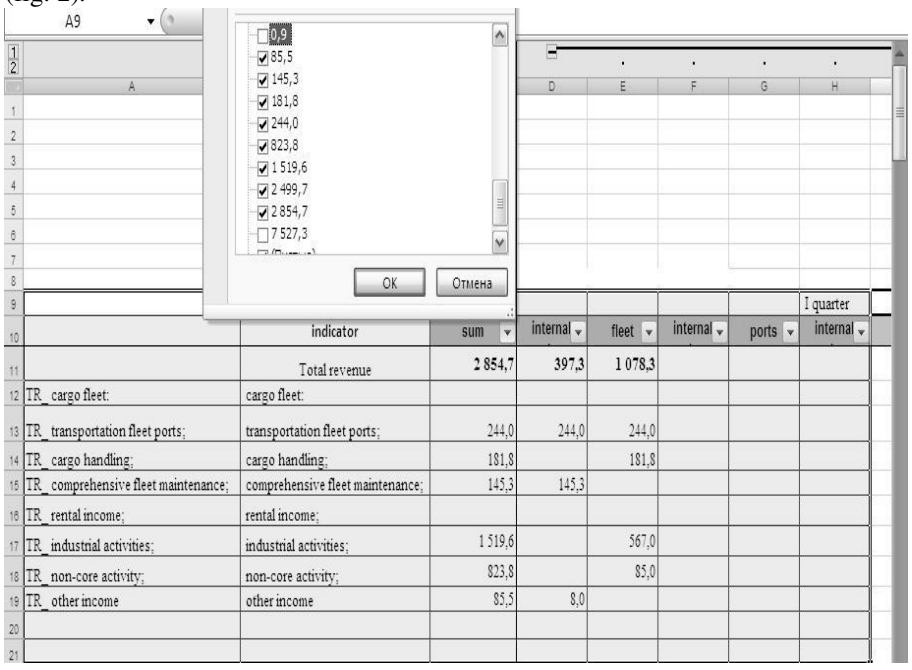


Fig. 1. Detailed fixed costs of the Kherson river port.

The model can not only be used for determining EBITDA and making a management reporting form, but can be a basis for implementation of the process of " Data Mining". Therewith, the spectrum of the problem solving by methods of Data Mining can be broad enough, from the sales revenue classification by types and business units to feasibility of a business unit in view of the internal business volume (fig. 2).



	indicator	sum	internal	fleet	internal	ports	I quarter internal
Total revenue		2 854,7	397,3	1 078,3			
TR_ cargo fleet:	cargo fleet:						
TR_ transportation fleet ports;	transportation fleet ports;	244,0	244,0	244,0			
TR_ cargo handling;	cargo handling;	181,8		181,8			
TR_ comprehensive fleet maintenance;	comprehensive fleet maintenance;	145,3	145,3				
TR_ rental income;	rental income;						
TR_ industrial activities;	industrial activities;	1 519,6		567,0			
TR_ non-core activity;	non-core activity;	823,8		85,0			
TR_ other income	other income	85,5	8,0				

Fig. 2. Detailed total revenue by origin and business-units including internal turnover.

6 Conclusion

Analysis of practical experience in the construction and operation of management accounting at the company in question, the reporting procedure for management and the algorithm for determining EBITDA are indicative of using the accounting system for the absorption - costing system which is focused on the owners (investors) requests regarding the effectiveness of the funds invested.

Organization of accounting using “1C Accounting for Ukraine" allows representing accounting as an information system in the form of a multidimensional data model to achieve a number of results, namely the creation of a single integrated accounting system, which meets both, "standard and non-standard" user requests; bridges the gap between the formation of actual financial and management accounting data; summarizes data for the preparation of management reporting forms with a given level of detail.

Use of the Pivot Tables in MS Excel provide for the appearance of new aspects of actual data usage, introduction of new connections between the data of financial and management accounting, which in turn does not lead to reconstruction of the whole accounting model and accounting database in general. In the pivot tables there are tools of data analysis allowing for the intellectual assessment, that is to summarize, group, delete unnecessary data, or increase the reliability by establishing links and accuracy of calculations. Good design of the tables can significantly facilitate the laborious process of making the management reporting forms and analyzing the company’s activities.

References

1. Kolesov, A. V.: Conceptual model of the analysis of expenses when using process approach. Vopr. economy and rights 12, 257--264 (2011)
2. Khalfallah, M., Moschetto, B. L., Teulon, F.: Evaluation of the profitability of companies financed by venture capital (CVC) listed on the French Market. Journal of Applied Business Research (JABR) 2, 313--328 (2014)
3. Strnadova, M., Karas M.: The Effect of Ownership Structure on the Performance of Manufacturing Companies. European Financial Systems, 588--595 (2014)
4. Law of Ukraine "On Accounting and Financial Reporting in Ukraine" № 996 XIV 16.07.1999 http://zakon2.rada.gov.ua/laws/show/996-14
5. Yatsenko, V.: Accounting and analytical system of multisegment enterprise: theoretical basis and practical implementation. Accounting and Auditing 11, 25--37(2014)

Statistical Analysis of Indexes of Capitalization of the Ukrainian Firms: an Empirical Research

Anastasiia Kolesnyk[1] and Ihor Lukianov[2]

[1]Kherson State University, 27, 40-Rokiv Zhovtnya Str., Kherson 73000, Ukraine
anastacia_kolesnik95@mail.ru
[2]Taras Shevchenko National University of Kyiv, 60, Volodymyrska Str., Kyiv 01601, Ukraine
lukia2007@ukr.net

Abstract. The document considers the performance and effectiveness of Ukrainian companies on the Warsaw Stock Exchange. With this end in view, the document examines the following issues: raising capital for investment, eliminating barriers for the Polish investors, capitalizing product reputation on the market, increasing investor recognition of companies, enhancing the corporate image. The document also analyses aspects of familiarity with local financial community, commitment to corporate governance standards, and the possibility of M&A. These figures are to indicate the status of Ukrainian companies on the Warsaw Stock Exchange.

Keywords. IPO, Warsaw Stock Exchange, Wig-Ukraine.

Key Terms. Industry, Management, Market, MathematicalModel, Research.

1. Introduction

Warsaw Stock Exchange is the largest national financial instruments exchange in the region of Central and Eastern Europe and one of the fastest-growing exchanges in Europe. The Group offers a wide range of products and services within its trading markets of equity, derivative, debt and structured products, electricity, natural gas, property rights, as well as clearing of transactions, operation of the Register of Certificates of Origin of electricity and sale of market data.

WSE started operation in 1991 as a company held 100% by the State Treasury. In 2010, the State Treasury arranged a public offering of WSE shares; as a result, shares of the Exchange were newly listed on the WSE Main Market on 9 November 2010. For instance, there are 471 companies represented at the WSE, including 51 foreign companies. Total market value of all companies is about 290 bl. Euro.

Poland's stock exchange market is growing stronger and becomes more international day by day. Its evolution is supported by the active marketing policy of the Warsaw Stock Exchange working to promote the entire infrastructure of Poland's capital market. These efforts have produced tangible results.

The Warsaw Stock Exchange conducts trading in financial instruments on three markets:

The Main List has been in operation since 16 April 1991. This market is supervised by the Polish Financial Supervision Authority and notified to the European Commission as a regulated market. The following securities and financial instruments are traded here: equities, bonds, pre-emptive rights, rights to shares, investment certificates, structured instruments, ETF and derivatives, i.e. futures contracts, options and index participation units.

- NewConnect is a market organised and maintained by the WSE as an alternative trading system. It was designed for startups and developing companies, especially from the sector of new technologies. NewConnect was launched on 30 August 2007. Instruments which may be traded under this alternative trading system include equities, rights to shares, pre-emptive rights, depository receipts, as well as other equity based instruments.
- Catalyst is a debt instruments market for municipal, corporate and mortgage bonds Founded on 30 September 2009, it consists of two trading platforms organised by the WSE as a regulated market and as an alternative trading system (ATS) for retail customers, and two analogous markets operated by BondSpot and designed for wholesale clients.

2. Problem Statements

If a company wants to be listed on the stock exchange, it should complete Initial public offering (IPO).[1]

Initial public offering (IPO) or stock market launch is a type of public offering in which shares of stock in a company usually are sold to institutional investors (that price the company receives from the institutional investors is the IPO price) that in turn sell to the general public, on a securities exchange, for the first time.

IPO benefits:

- Access to capital to fund growth

Public placement of shares on a stock exchange allows the company to attract capital to fund both organic growth (modernization and upgrade of production facilities, implementation of capital-intensive projects) and acquisitive expansion. If retained earnings and debt funding are insufficient, IPO becomes one of the most realistic and convenient ways to secure the continuing growth of the business. It provides access to a massive, timeless pool of capital and boosts the investment credibility of the business.

- Creation of liquidity and potential exit for the current owners

Formation of a public market for the company's shares at fair price creates liquidity and provides an opportunity to sell the shares promptly with minimal transactional costs. The private owners of the company can dispose of their stakes in the business both during an IPO (this route is often taken by the minority financial

[1] IPO calendar, http://www.fixygen.ua/calendar/ipo/

investors such as venture or private capital funds) and at a later stage (this is often preferred by the majority shareholders).

- Maximum value of the company

Normally, an IPO is an offer to a large number of institutional and retail investors to become shareholders of the company. The very multitude of large investors and their confidence in the liquidity of their investment in a public entity assure the current owners of a private company about achieving the maximum possible valuation of the business at the time of an IPO or afterwards.

- Enhancement of the company's public profile

Listing on a recognized stock exchange means that the business will receive wide media coverage, usually a very favorable one, thus increasing the company's visibility and recognition of its products and services. The company's activities will also be reflected in the reports by professional financial analysts. Such public profile supports liquidity of the shares and contributes to the expansion of the business contacts. It also helps to increase confidence among the company's business partners.

- Improvement in debt finance terms

For domestic (Ukrainian or other CIS country-based) financial institutions – used to working with the low-transparency businesses and often inadequate financial reporting – a company listed on a recognized stock exchange becomes a desirable and reliable partner. Banks are often ready to extend loans to public companies in larger amounts, under smaller collateral, for longer maturities and with lower interest rates. Even the largest and most prestigious banking institutions are keen to work with public companies – whose transparency and corporate governance serve as additional factors of confidence for banks and other suppliers of credit.

- Extra assurances for partners, suppliers and clients

Partners and contractors of a public company feel more confident about its financial state and organizational capabilities as compared to those of a non-transparent private business. Partners take additional comfort in the fact that the public company has gone through rigorous legal, financial and corporate due diligences – all of which are required for a successful completion of an IPO. Confidence among partners and contractors is a sound foundation for stable and predictable business relations with the public company, and allows the latter to obtain additional leverage in negotiating better terms for doing business.

- Enhanced loyalty of key personnel

Publicly available information about the share price of a public company allows development of employee motivation schemes based on partial remuneration of staff in the form of participation in the equity capital (for example, share options). Equity-based incentive schemes stimulate the key personnel to become more efficient in their work in order to support the company's growth rates and profitable development – which in turn increase the operational and financial efficiency of the company and its market value.

- Superior efficiency of the business

Conduct of various due diligences during the IPO process requires a thorough and comprehensive analysis of the company's business model. During the IPO implementation process, certain internal changes take place, including modification of

the organizational structure; selection of the key personnel and delegation of responsibilities; improvement of internal reporting and controls; as well as critical evaluation of the efficiency of the entire business. Normally, such extensive internal efforts result in significant improvements of the communication system, management and controls; they also help eliminate any previously hidden shortcomings in the internal functioning of the business.

The IPO process can be very complicated. There are certain steps you must take along the way. These steps will help insure that your IPO is successful.

- Planning for the IPO Process

You need to determine at the beginning whether it's a good time for an IPO. Choosing the ideal time to go public is very important. Plan in great detail what you hope to accomplish. Examine your financial needs and wants.

It's helpful for a business to act like a public company even before it goes public. This can be done a couple of years in advance of the IPO. Develop a business plan and prepare financial statements.

- Choosing Underwriters

Most companies use underwriters to help them with IPOs. Choosing the right underwriters is key to having a successful offering. They're usually the ones responsible for buying and selling the securities to the public. They're also responsible for investigating your business to verify the financial information given to the investors. You should select the underwriters at least a few months before the IPO date.

- Filing a Prospectus

Your business must file a registration statement with the US Securities and Exchange Commission (SEC). This statement contains detailed information about the offering. It also includes information about the business, its financial history and its future plans.

The registration statement becomes the preliminary prospectus once it's filed with the SEC. A prospectus is a legal document explaining the securities offered to the public. The preliminary prospectus is also called the red herring. It's called this because red ink is used on the front page to indicate certain information may change.

The SEC will examine the registration statement during a "cooling off" period. It informs the business of any necessary changes. The statement becomes the official prospectus once any necessary amendments are made. The prospectus can be used by the public to help them determine whether they want to purchase the securities for sale.

- IPO Promotion

A business going public has to market the IPO. Representatives from the company and underwriters go on a "road show" around the country. They make numerous presentations to potential investors. Typical stops include New York, Chicago, Boston, Los Angeles and San Francisco. Even international trips may be set up for overseas investors.

- Final Offering Price and Amount

Choosing the final offering price and the amount of securities to be sold are very important decisions. Market conditions and the expected demand for the securities

need to be examined closely. These final decisions are usually made right before the offering.

- Selling on the Stock Market

The IPO is normally declared effective a few days after the final prospectus is received by the potential investors. This declaration is usually done after the stock market has closed. The securities will then be available for trade the next day. The IPO will hopefully be successful and provide new capital for the business for their present and future plans.

Taking into consideration the benefits of being involved in the Warsaw Stock Exchange activities, many Ukrainian companies want to be listed. However, in order to be listed at the WSE, a company is obliged to meet special requirements as follows:

1. Only a joint stock company may be an issuer of shares listed on the WSE. This does not bar entities operating under any other legal form from listing, but their owners need to transform them into joint stock companies or establish joint stock companies and transfer the entities assets thereto.

2. As a next step, the General Shareholders Meeting should adopt a resolution approving a public offer of shares and an application for admission of the shares to trading on the regulated market.

3. The decision to apply for admission to trading in the regulated market may require the preparation of a relevant information document (issue prospectus or information memorandum) so the company will need to work with:

- an auditor who will audit the company's financial statements and convert them into a format comparable year to year;
- a brokerage house which will offer the company's shares in a public offer.

Depending on the issuer's individual needs, the company may need to hire legal and financial advisors. The contents of the issue prospectus are laid down in the Commission Regulation (EC) No. 809/2004 of 29 April 2004 as regards information contained in prospectuses.

4. Next, the company will need to submit the working draft of the issue prospectus to the Polish Financial Supervision Authority (KNF). KNF may communicate its comments, and once the company has accommodated those in the final draft of the issue prospectus, KNF will decide whether to approve the prospectus.

5. Before opening the public offer, the issuer will need to execute an agreement with the National Depository for Securities (KDPW) whereby the securities subject to the public offer will be registered by the Depository.

6. The public offer may now proceed. Before allocated shares of a new issue are registered, rights to shares may be traded on the WSE.[2]

7. Once the offer is closed, the company will submit an application for the admission of shares (and possibly also rights to shares) to stock exchange trading on the main or the parallel market. The WSE Management Board will examine the application. The application must include, among others, the final draft of the issue prospectus accommodating all recommendations made by the KNF.

[2] WSE, http://www.ipowse.com.ua/about/

8. Once all shares introduced to trading are deposited with KDPW, the public offer is closed, and the shares of the new issue registered by the court, the company will file with the WSE Management Board an application for the introduction of shares to trading on the main or the parallel market. The WSE Management Board will indicate the trading system and the date of the first trading session.

Some Ukrainian companies have already been listed at the regular market of the WSE. There are some descriptions of such companies:

Kernel is a leading diversified agribusiness company in the Black Sea region listed on the Warsaw Stock Exchange. Handling about 6 million tons of agricultural commodities per year, Kernel supplies international markets with grain and sunflower oil produced in Ukraine and Russia. The production assets extend from black soil farmland to oilseed crushing plants supported with essential agricultural infrastructure including silos and deep-water export terminals. 2007 marked also a new stage in the development as Kernel became a publicly listed company: listed Kernel on the Warsaw Stock Exchange in November 2007 and new shareholders entered the capital of the Company to participate in the growth story.[3]

Founded in 1993, "Astarta–Kyiv" is a vertically integrated agro-industrial holding specializing in sugar and agricultural production. It has proven to be a growing, transparent company, as well as a reliable partner and supplier. Implementing a strategy of vertical integration, ASTARTA created a fully integrated production cycle of sugar from growing the beet to sugar production and sales. Growing sugar beets lowers dependence on the external supply of sugar beets, lowers the cost of produced sugar, and guarantees constant manufacturing and the highest possible yield and quality. In August 2006, ASTARTA's shares are listed on the Warsaw Stock Exchange.

Group of Companies «Ovostar Union» is one of Ukrainian leading agro-industrial companies, entering TOP 3 Ukrainian egg producers. The history maintains 14 years of experience, leadership and innovations. The main advantage is vertically integrated business organization structure, providing accurate product quality control at any production stage. Each enterprise of GC «Ovostar Union» is an integral part of whole business, important and high-grade link, performing its obligations effectively, and in this way ensuring the common great result. It's the conformity of all Company activity vectors determining products guaranteed quality and in general high business profits. In 2011 GC «Ovostar Union» has debuted on the Warsaw Stock Exchange and attracted 93 million zlotys in an initial public offering (IPO). It has been using to carry out the group's investment program. 1.5 million shares were placed (25% of the capital) at a price of 62 zlotys per share. The price is equal to the maximum price.[4]

Industrial Milk Company (IMC) is an integrated agricultural business operating in Ukraine. In May 2011 IMC conducted IPO on Warsaw Stock Exchange.

The main areas of IMC's activities are:

- cultivation of grain & oilseeds crops, potato production
- storage and processing of grain & oilseeds crops

[3] Kernel, http://www.kernel.ua/en/

[4] Ovostar, http://www.ovostar.ua/ru/

- dairy farming

IMC is among Ukraine's top-10 agricultural companies (source of ratings: AgriSurvey, "The largest agro holdings in Ukraine", 2014, based on results of year 2013). In May, 2011 the company completed IPO on Warsaw Stock Excange, IMC raised US$ 24,4 mln to finance the development of the company.[5]

Milkiland is an international diversified dairy producer with the core operations in the CIS and EU. The Group's total annual milk processing capacity exceeds 1 million tons. The company is proud to produce natural dairy from the best milk: a wide range of fresh dairy, different types of cheese, and butter to satisfy the consumers in their everyday needs for healthy and tasty foods. Milkiland's dry dairy products are exported to over 30 countries. The international Dobryana brand is popular among cheese and dairy consumers in Ukraine, Russia and other CIS countries. Ostankinskoye is a traditional brand for whole-milk products produced by Ostankino Diary Combine, well known by Moscow consumers. Fresh dairy under Ostrowia brand is also well known in Poland. 21.48% of the shares of Milkiland N.V. are in the free float at the Warsaw Stock Exchange.[6]

KDM Shipping is one of the leaders of the Ukrainian shipping industry, primarily involved in the niche segment of dry bulk river-sea freight in the Black, Azov and Mediterranean Sea regions. The Group's cargo fleet consists of 10 river-sea, dry cargo vessels of total 29,673 DWT, which due to their shallow draft can access major river and sea ports in Black and Azov Sea regions. The Group also provides passenger river transport services in the Kiev and other regions of Ukraine (operating the fleet of 8 passenger river vessels), as well as ship repair services at its own shipyard located in the city of Kherson. According to International Economic Rating "League of the Bests" in 2012, the Group is ranked as # 1 in river activity and 3rd in maritime activity in Ukraine.[7] The Group has developed a vertically integrated business model. The Group's main activity of dry-bulk shipping is supported by its own ship repair yard, its own ship agency in selected ports as well as its own crewing department. Such business model allows the Group to benefit from certain cost efficiencies and sustain competitive advantages.

Coal Energy S.A. was incorporated in the Grand Duchy of Luxembourg and has been listed on the Warsaw Stock Exchange in Poland since 08 August 2011. Coal Energy S.A. is a holding company for a group of 12 companies operating in the coal industry in Ukraine which rank the Company as the 3rd largest in terms of coal deposits and the 4th largest in terms of extraction volumes private coal mining enterprise in the country based on the calendar year 2011 data (hereinafter – "the Group" and/or "Coal Energy").

Principal business activities of Coal Energy are mining, beneficiation and sales of thermal and coking coals as well as dual purposes coal. Companies of the Group are directly cooperating with all the largest heat and power generation plants and metallurgic plants in Ukraine. Due to its favorable geographical location and wide

[5] Industrial Milk Company, http://www.imcagro.com.ua/ru/
[6] Milkiland, http://www.milkiland.nl/ru
[7] KDM Shipping, http://kdmshipping.com/

products' palette Coal Energy is able to export produced thermal coal to power producing stations in Turkey, Moldova, Bulgaria, Slovakia and other countries where the Group has established contacts with the largest power generations.

KSG Agro is one of the most dynamically developing agricultural groups in Ukraine. We take innovative approach for our key business philosophy. The search of non-standard approaches and creative decisions requires professional experience and passion for a native land. Our main values are people and land. Everything we produce is made by people and for people. That is why we highly appreciate your trust and consider it a key factor of the Group's success. Trust is impossible without responsibility. Providing an example of long-term and prosperous cooperation among the state, society, and agricultural business, we create the foundation for leadership.

Nowadays the Group gained a lot to be proud of; however we still strive for more. Highly-profitable agricultural industry with a high level of diversification and vertical integration gives growth prospects and sense of security. We have got the main factor for this purpose which is our team of like-minded professionals. The core business of the Group is cultivation of land and production of agricultural crops. Complex approaches to farming and focus on intensive development of business ensure high profitability and also create the conditions facilitating high yields.

The Group focuses on the following business directions:

- Crop production
- Vegetables production
- Fruits production
- Food processing business
- Supplies of food to retail networks

3. Results

The full list of Ukrainian companies listed at the Warsaw Stock Exchange:

Table 1. Ukrainian companies at the WSE

Name of the company	Date of IPO	The percentage of capitalization	Sum of the capitalization
Astarta Kiev	August 2006	20% of shares	23,7 bl. Euro
Kernel Holding	November 2007	37,98% of shares	136,5 bl. Euro
Agroton public limited	November 2010	26,2% of shares	38,25 bl. Euro
Milikiland	December 2010	22,4% of shares	60 bl. Euro
Sadovaya Group	December 2010	25% of shares	23,15 bl. Euro
Industrial Milk Copmany	May 2011	23,9% of shares	20,7 bl. Euro
KSG Agro	May 2011	33% of shares	27 bl. Euro
Westa	June 2011	25% of shares	32 bl. Euro
Ovostar Union N.V	June 2011	25% of shares	22,2 bl. Euro

Coal Energy	August 2011	25% of shares	56 bl. Euro
KDM Shipping	August 2012	11 % of shares	6,3 bl. Euro

WIG-Ukraine is the second national index calculated by WSE. Its portfolio includes companies listed on the Warsaw Stock Exchange, where a company or head office is located in Ukraine, or whose business is conducted to the greatest extent in this country. It has been calculated since May 4, 2011. The historical values were recalculated since December 31, 2010 (the base date). The initial value of the index was 1000 points. WIG-Ukraine is a total return index and thus when it is calculated it accounts for both prices of underlying shares and dividend and subscription rights income.[8]

Composition of WIG-Ukraine
(as for 24 February 2015)

Table 2. The composition of WIG-Ukraine

Instrument	ISIN	Share	Market value of shares (PLN)	Quota (%)
KERNEL	LU0327357389	9,509,000	286,601,260	37.877
ASTARTA	NL0000686509	9,253,000	227,346,210	30.046
OVOSTAR	NL0009805613	1,725,000	122,233,500	16.154
IMCOMPANY	LU0607203980	9,809,000	63,169,960	8.348
MILKILAND	NL0009508712	8,276,000	26,152,160	3.456
KDMSHIPNG	CY0102492119	3,329,000	18,775,560	2.481
COALENERG	LU0646112838	11,252,000	6,638,680	0.877
KSGAGRO	LU0611262873	5,093,000	5,755,090	0.761

$$WIG - Ukraine = \frac{M(t)}{M(0) * K(t)} * 100\% \quad (1)$$

M(t)-capitalisation of index portfolio at session "t"
M(0)-capitalisation of index portfolio at base date
K(t)-adjustment coefficient for session "t"

Now, using application package of MS Excel, we shall determine the effect of Ukrainian companies to index. The following table draws up the results:

[8] Wig-Ukraine, http://www.gpw.pl/opis_indeksu_WIG-Ukraine_ru

Table 3. The effect of Ukrainian companies to index

b8	**b7**	**b6**	**b5**	**b4**	**b3**	**b2**	**b1**	**b0**
1.47	-1.70	0.889	2.69	1.38	0.385	2.69	4.96	6.08
0.15	0.996	0.042	0.22	0.159	0.052	0.028	0.039	3.42
t(b8)	t(b7)	t(b6)	t(b5)	t(b4)	t(b3)	t(b2)	t(b1)	t(b0)
9.99	1.71	21.01	12.26	8.69	7.33	94.799	126.43	1.78
Signi-ficant	Insigni-ficant	**Signi-ficant**	**Signi-ficant**	**Signi-ficant**	**Signi-ficant**	**Signi-ficant**	**Signi-ficant**	Insigni-ficant
tkr	1.97							

Obtained results in the table are equal to the following equation multiple linear regressions:

y=6.079+4.956*X1+2.686*X2+0.385*X3+1.379*X4+2.693*X5+0.889*X6-1.705*X7+1.473*X8 (2)

We shall interpret it as following:

- 6.079 – performs, that after zero changes of prices of shares of all companies індекс зросте the index shall be increased by 6%
- 4.956– performs, that after 1 % increase of prices of shares of Kernel Holding the index shall be increased by 5%
- 2.686– performs, that after 1 % increase of prices of shares of Astarta Kiev the index shall be increased by 2,7%
- 0.385– performs, that after 1 % increase of prices of shares of Ovostar Union the index shall be increased by 0,38%
- 1.379– performs, that after 1 % increase of prices of shares of Industrial Milk Co. the index shall be increased by 1,34%
- 2.693– performs, that after 1 % increase of prices of shares of Milkiland the index shall be increased by 2,7%
- 0.889– performs, that after 1 % increase of prices of shares of KDM Shipping Public Ltd the index shall be increased by 0,9%
- 1.705– performs, that after 1 % increase of prices of shares of Coal Energy the index shall be increased by 1,7%
- 1.473– performs, that after 1 % increase of prices of shares of KSG Agro the index shall be increased by 1,5%

Application of Student's t-Tests shows, that almost all companies have an effect on the index. Though, we can point out two of them**,** namely Kernel Holding, Astarta Kiev**,** having the most considerable effect on the index. Their shares account for the most considerable percentage, in particular for 67% of Wig-Ukraine index. More importantly, Kernel Holding and Astarta Kiev are the first companies to be listed at the regular market. They have already adapted to the market, having a great superiority as compared to other Ukrainian companies. Moreover, Kernel Holding is also included in composition of WIG30 index which is based on the value of portfolio of 30 major and most liquid companies on the WSE Main List. Both companies are engaged in ag-

ricultural business, which is the most distinctive feature of the Ukrainian companies, involved in international business activity.

4. Conclusion

The following results were obtained in the course of this research:

- The primary reason (motive) for the Ukrainian companies to be listed at WSE is the possibility of acquiring access to the capital;
- The reputation and credibility of the stock company (market) provides the Ukrainian companies with the possibility of finding new investors or potential financial partners;
- As of today, there are 11 Ukrainian companies at WSE and 2 companies at the NewConnect market;
- More than two billions Euros were attracted by the Ukrainian companies;
- In fact, Warsaw Stock Exchange is the most effective way of entry into the international market for the Ukrainian companies, which, in the meantime, enables the companies to strengthen their positions on the market, raise capital and gain investments in the future.

References

1. Asprem, M. Stock Prices, Asset Portfolios and Macroeconomic Variables in Ten European Countries, Journal of Banking and Finance 13, 589–612. (1989)
2. Baker, M. and Wurgler, J. Investor Sentiment in the Stock Market, Journal of Economic Perspectives 21, 129-151. (2006)
3. Chen, F., Roll, R. and Ross, S. Economic Forces and the Stock Market, Journal of Business 59, 383–403. (1986)
4. Dopke, J., Hartmann, D. and Pierdzioch, C. Forecasting Stock Market Volatility with Macroeconomic Variables in Real Time, Discussion Paper, Deutsche Bundesbank. (2006)
5. Errunza, V. and Hogan, K. Macroeconomic Determinants of European Stock Market Volatility, European Financial Management 4, 361-377. (1998)
6. Garcia, V. and Liu, L. Macroeconomic Determinants of Stock Market Development, Journal of Applied Economics 2, 29-59. (1999)
7. Homa, K. and Jaffee, D. The Supply of Money and Common Stock Prices, The Journal of Finance 26, 1045-1066. (1971)
8. Thalassinos, E.I., Thalassinos, P.E. Stock Markets' Integration Analysis, European Research Studies, Vol. IX, Issue 3-4. (2006)

The Hybrid Service Model of Electronic Resources Access in the Cloud-Based Learning Environment

Mariya Shyshkina

Institute of Information Technologies and Learning Tools of the National Academy of Pedagogical Sciences of Ukraine, Berlinskii Str., 9, Kyiv, Ukraine
marple@ukr.net

Abstract. Nowadays, the search for innovative technological solutions to the organization of access to electronic learning resources in the university and their configuration within the environment to fit the needs of users and to improve learning outcomes has become key issues. These solutions are based on the emerging tools among which cloud computing and ICT outsourcing have become very promising and important trends in research. The problems of providing access to electronic learning resources on the basis of cloud computing are the focus of the article. The article outlines the conceptual framework of the study by reviewing existing approaches and models for the cloud-based learning environment's architecture and design, including its advantages and disadvantages, and the features of its pedagogical application and the experience of it. The hybrid service model of access to learning resources within the university environment is described and proved. An empirical estimation of the proposed approach and current developments in its implementation are provided.

Keywords: hybrid model, learning environment, cloud computing, university.

Key Terms: ICTInfrastructure, Model, TeachingProcess

1 Introduction

Progress in the area of ICT and network technology has provided new insights into the problems of the formation and development of the educational environment of the university, showing a need for advanced ICT access, especially with regard to the use of the cloud-based tools and resources. There is a need for modernization of learning technologies, supported by emerging ICT, on the basis of advanced network infrastructures.

Cloud computing technology (CC) enhances multiple access and joint use of educational resources at different levels and domains. On the basis of this technology, it is possible to combine the corporate resources of the university within a united framework. To achieve this aim, a set of cloud-based learning models should be created for the elaboration and design of learning resources and the learning environment architecture to deliver access to learning resources.

The purpose of the article is analyse the current trends in the university cloud-based learning environment formation from the perspective of different service models used, and to substantiate and validate the hybrid service model of access to the learning resources.

The *research method* involved analysing the current research (including the domestic and foreign experience of the application of cloud-based learning services to reveal the concept of the investigation), examining existing models and approaches, estimating the current state of research development, considering existing technological solutions and psychological and pedagogical assumptions about better ways of introducing innovative technology, and conducting pedagogical experiments, surveys and expert evaluations.

2 Problem Statement

The progress of ICT has a significant impact on the formation of the educational environment of the university bringing with it new models of the organization of learning activity which arise on the basis of decisions about innovative technology. In this regard, the phenomenon of the cloud-based learning environment has come to the forefront as it has many progressive features including better adaptability and mobility, as well as full-scale interactivity, free network access, a unified infrastructure among others [4, 19, 20].

The challenges of making the information technology infrastructure of the university setting fit the needs of its users, taking maximum advantage of modern network technologies, and ensuring the best pedagogical outcomes to increase the learning results, has led to the search for the most reasonable ways of organizing tools and services within the framework of this environment. For this purpose, the modelling and analysis of its structure and functions, and determining the possible types and forms of learning activity in the organization have come to the fore. Among the priority issues for ICT infrastructure design is the access to software and electronic educational resources provision [4]. To choose the best solution there is a need to consider existing approaches and models to reveal possible ways of service deployment, and to analyse the existing experience of its use.

3 State of the Art

According to the recent research [4, 9, 13, 18, 19], the problems of implementing cloud technologies in educational institutions so as to provide software access, support collaborative learning, implement scientific and educational activities, support research and project development, exchange experience and are especially challenging. The formation of the cloud-based learning environment is recognized as a priority by the international educational community [16], and is now being intensively developed in different areas of education, including mathematics and engineering [2, 8, 11, 25, 27].

The transformation of the modern educational environment of the university by the use of the cloud-based services and cloud computing (CC) delivery platforms is an important trend in research. The topics of software virtualization and the forming of a unified ICT infrastructure on the basis of CC have become increasingly popular lines of research [8, 18, 23]. The problems with the use of private and public cloud services, their advantages and disadvantages, perspectives on their application, and targets and implementation strategies are within the spectrum of this research [7, 8, 25].

There is a gradual shift towards the outsourcing of ICT services that is likely to provide more flexible, powerful and high-quality educational services and resources [4]. There is a tendency towards the increasing use of the software-as-a-service (SaaS) tool. Along with SaaS the network design and operation, security operations, desktop computing support, datacentre provision and other services are increasingly being outsourced as well. Indeed, the use of the outsourcing mechanism for a non-core activity of any organization, as the recent surveys have observed happening in business, is now being extended into the education sector [9]. So, the study of the best practices in the use of cloud services in an educational environment, the analysis and evaluation of possible ways of development, and service quality estimation in this context have to be considered.

The valuable experience of the Massachusetts institute of technology (MIT) should be noted in concern to the cloud based learning environment formation in particular as for access to mathematical software. The Math software is available in the corporate cloud of the University for the most popular packages such as *Mathematica, Mathlab, Maple, R, Maxima* [27]. This software is delivered in the distributed mode on-line through the corporate access point. This is to save on license pay and also on computing facilities. The mathematics applications require powerful processing so it is advisable to use it in the cloud. On the other case the market need in such tools inspires its supply by the SaaS model. This is evidenced by the emergence of the cloud versions for such products as Sage MathCloud, Maple Net, MATLAB web-server, WebMathematica, Calculation Laboratory and others [2, 8]. Really there is a shift toward the cloud-based models as from the side of educational and scientific community, and also from the side of product suppliers. The learning software actually becomes a service in any case, let it be a public or a corporate cloud.

There are many disciplines where it is necessary to outsource the processing capacity: for example, the computer design for handling vast amounts of data for graphics or video applications. This is also a useful tool used to support the collaborative work of developers, as the modern graphical applications appear to be super-powerful and require joint efforts [7]. There is a research trend connected to the virtual computing laboratories (VCL) [14, 26] delivered in the cloud-based paradigm. This trend is inherent in the field of informatics, and learning resources for processing and sharing are needed.

Nowadays there are various universal cloud consumer applications, in particular MicrosoftOffice 365, Google Docs and others which gain an appropriate use in educational process [9, 23]. There is also a wide range of cloud services such as

online photo and video editors, web pages processors, services for translation, check spelling, anti plagiarism and many others which are now available [23].

There is a principal transformation of approaches in concern to services supply within the cloud based infrastructure. It is considered to be a new stage of the service oriented models development [10, 24]. There is a branch of research devoted to the service oriented infrastructure in this actual perspective. The issues of service oriented architecture development and are described in [10]. The problem of turning software into a service is also posed [24]. For example, more powerful approaches for services integration appear while services compositions are used as building blocks in a process of elaboration of programming code [14]. The CC development brought the term the *service orchestration* into scientific discussion while number of web services can be combined to perform the higher level business process to manage and coordinate execution of the component processes [12]. In this regard the notion of the global software development (GSD) is considered as novel trends overcome geographical limits [12]. There is a significant revise of approaches to ICT services elaboration and this is concerned to its integration and composition.

An essential feature of the cloud computing conception is dynamical supply of computing resources, software and hardware its flexible configuration according to user needs. Due to this approach, access to educational software set on a cloud server or in a public cloud is organized. So comparison of different approaches and cloud models of software access is the current subject matter of educational research [7, 8, 23, 25]. Also the problems of quality criteria for software choice in the learning complexes to be implemented in a cloud arise. Despite of the fact that the sphere of CC is rather emerging there is a need of some comparison of the achieved experience to consider future prospects [23].

Another set of problems is concerned with the hybrid service models and infrastructure solutions combining different public and corporate services on the united platform. This trend is now especially promising for the sphere of education [8, 17]. The challenge regarding novel technological solutions and their impact guide the search for the most reasonable method of implementation.

Thus, in view of the current tendencies, the research questions are: how can we take maximum advantage of modern network technologies and compose the tools and services of the learning environment to achieve better results? What are the best ways to access electronic resources if the environment is designed mainly and essentially on the basis of CC? This brings the problem of cloud-based services modelling, integration and design to the forefront.

4 Pedagogical Aspects of Electronic Resources Delivery and Indicators of Research

Cloud computing technology is now one of the leading trends in the formation of the information society. It constitutes an innovative learning concept and its implementation significantly affects the content and form of different types of activities in the sphere of education [4, 13, 18].

Along with the emergence of cloud computing, the number of objects, developments and domain applications are continually growing, which indicates the rapid spread of the innovation [20]. The concept of *the cloud-based learning environment* is now in line with the wider trend; that is to say, the ICT environment of the university, where some didactic functions as well as some fundamentally important functions of scientific research are supported by the appropriately coordinated and integrated use of cloud services [20]. The *aim* of the cloud-based learning environment formation is to meet the users' educational needs. To do this, the introduction of cloud technology in the learning process should to be holistic and carried out according to the principles of *open education*, including meeting the following needs: the mobility of students and teachers, equal access to educational systems, providing qualitative education, and forming and structuring of educational services [3, 20].

The main elements of the cloud computing conception, including the types, application service models, essential features, ICT architecture and others, are reflected in the structure of the modern educational organizational systems [5]. Therefore, a number of concepts and principles that characterize the development and application of CC-based services are significant in the consideration of the educational environment design.

The concept of *electronic educational (learning) resources* (EER) appears to be the centre of attention. In particular, at the Institute of Information Technologies and Learning Tools of the National Academy of Pedagogical Sciences of Ukraine the conception that provided the definition of electronic educational resources (EER) its classification, and the ways it can be applied has been developed and proposed [5].

According to the definition given in [5, p.3], "The electronic educational resources are a kind of educational tool (for training, etc.) that are electronically placed and served in educational system data storage devices which are a set of electronic information objects (documents, documented information and instructions, information materials, procedural models, etc.)".

The elaboration of the electronic learning resources should be considered as a specific activity, which is linked to the mandatory need to take into consideration the psychological and pedagogical aspects of building an educational system methodology, the design of an open computer-based learning environment, and the involvement of the scientific and pedagogical staff, including the best teachers and educators [4].

Cloud Service – is a service that makes software applications, data storage or computing capacity available to users over the Internet [16]. These services are used to supply the electronic educational resources that make up the substance of a cloud-based environment, and to provide the processes of elaboration and use of the educational services.

Electronic resources appear to be both the objects and the tools of activity for a learner; therefore, these resources are used to maintain certain functions that are realized in the learning process. By the *educational service* we mean a service provided at the request (in response to an inquiry etc.) of a user that meets some

service function carried out by the organization or institution (service provider, outsourcer) [4].

Nowadays, the various types of electronic educational resources that may be delivered by the cloud in the learning environment of the university constitute libraries and depositories of EER or those retrieved when open analytical information systems are used. The EER supports different types of learning and research activities, such as theoretical material studies, the search for useful information, translation and grammar checking, task solutions, testing, training, simulation, making experiments and others.

Along with the development of information and communication technologies for education, the ways and tools of access to electronic resources have changed in an evolutionarily way and its custom properties have improved. There are new types of EER supplied by means of cloud technologies. The EER of the public cloud can occupy the role of software for general purposes such as office applications, systems support processes for communication and data exchange and others, and also the special software designed for educational use [13, 23]. The number of EERs is increasing and this trend is likely to intensify. By means of CC-based tools, a significant lifting of restrictions on the implementation of access to qualitative leaning resources may be achieved. Now, these questions are not a matter of future perspective, they need practical implementation. For this purpose, the problem of the design and delivery of electronic educational resources in the cloud-based environment is a complex one and not only should technological needs be considered, but also the pedagogical aspects.

With the advance of ICT, CC technologies appear to be a factor in the change in the content, methods and organizational forms of learning and development of the open education models. Now, cloud computing technology is used to improve the educational process through the presentation of a modern learning content adequate to the goals set, the quality monitoring and evaluation of learning results at the various stages, the creation of new organizational forms of learning, the creation of innovative educational and scientific resources and electronic systems and their implementation in the process of students' self-study and classroom study, advances in computer-aided and mixed models of training and so on [20].

As noted in [5], the necessary measures for the development of the human resources' component of the software industry created in Ukraine that concern the organization of EER access in the educational institutions are as follows: improvement to EER quality, scientific-methodological research on the implementation of innovative technologies and prospective models and methods in education, the development of the normative regulatory framework, strengthening the firms and companies in the IT industry and their participation in providing educational hardware and software and so on.

Due to the significant educational potential and novel approaches to environmental design, its formation and development, these questions remain the matter of theoretical and experimental studies, the refinement of approaches, and the search for models, methods and techniques, as well as possible ways of implementation [4].

To carry out research and experimental activities and the implementation and dissemination of the results, the Joint research laboratory of the Institute of Information Technologies and Learning Tools of the NAPS of Ukraine and the Kherson State University was created in 2011 with the focus on issues of educational quality management using ICT [29].

As part of the programme of joint research work, the Kherson State University was approved as an experimental base for research on the definition and experimental verification of the didactic requirements and methods of evaluating the quality of electronic learning resources in the educational processes of the pilot schools [29]. The purpose of the experiment carried out was to identify and experimentally verify the requirements and methods of evaluating the quality of the electronic learning resources used in the educational process in secondary schools [29].

The quality evaluation of EER in the cloud-based learning environment is a separate line of work in the Laboratory's research. In this case, there are different approaches and indicators. The access organization has been changed so the models of learning activity have been changed also. There are the following questions: What features and properties have to be checked so as to measure the pedagogical effect of the cloud-based approach? With regard to the pedagogical innovation, what are the factors influencing pedagogical systems, their structure and organization? Is the improvement in learning results achieved due to the cloud-based models? In this context, the quality of EER is a criterion for estimating the level of organization and functioning of the cloud-based learning environment.

With regard to this, the following *hypothesis* is to be posed: the design of the learning environment on the basis of cloud models of access to learning resources contributes to the improvement of the quality of these resources and the improvement of the processes in this environment and their organization and functioning, resulting in an improvement in learning results.

In the cloud-based learning environment, new ways of EER quality control arise. There are specific forms of the organization of learning activity related to quality estimation. For example there are e-learning systems based on the modelling and tracking of individual trajectories of each student's progress, knowledge level and further development [28]. This presupposes the adjustment, coordination of training, consideration of pace of training, diagnosis of achieved level of mastery of the material, consideration of a broad range of various facilities for learning to ensure suitability for a larger contingent of users. The vast data collections about the students' rates of learning are situated and processed in the cloud [28]. There are also collaborative forms of learning where the students and teachers take part in the process of resource elaboration and assessment; this is possible in particular by means of the SageMathCloud platform [2].

The prospective way of the estimation of the quality of learning resources is by means of the cloud-based environment. As the resources are collectively accessed, there is a way to allow experts into the learning process so they may observe and research their functioning. This is a way to make the process of quality estimation easier, more flexible and quicker. The process of estimation becomes anticipatory and

timely. The estimation may be obtained just once along with the process of EER elaboration, and it is very important to facilitate the process.

This method of estimation was developed and used in the Joint laboratory of EER quality control [29]. In this case, the different quality parameters will be detailed and selected. It is important that the psychological and pedagogical parameters are estimated in the experimental learning process, while the other types of parameter such as technological or ergonomic may be estimated out of this process.

The indicator of accessibility is also included in the focus of this investigation [15]. This property is essential because it is prior to other features such as scientific correctness, clarity, consistency and others, which may be researched only if this resource is available and feasible. The accessibility is characterized in turn by such features as convenience of the access organization, ease of use, interface consistency, advisability and others.

5 The Types of Service Models for Learning Resources Access

According to recent research, a *unified storage architecture* is an advantage of cloud based settings allowing application virtualization [18, 19]. This architecture is designed for the large complex data sets retrieving and management and it has the following features:

- different storage protocols are maintained in the same system (FC, NFS, FcoE, CIFS, iSCSI);
- various storage functions are implemented within the same device (storage, security, backup, recovery);
- storage space is scaled and modified without interruption of usual operations;
- data are integrated in a standard pool, which can be controlled over a network and managed via standard software package;
- data are used for different range of applications while storage area is not necessarily separated to enable saving computing capacity through virtualization.

Application virtualization is a technology for software access and development without installing it on a personal computer of a user. Data processing and storage is fulfilled in a data centre, and working with applications is not different for a customer from the working with applications installed on his (her) own desktop.

There are three main types of *service models* [16] that correspond to different ways the ICT outsourcing used to provide software and computing resources access [4]. In particular, *SaaS* (Software-as a Service) is to deliver software applications of a provider via the Internet; *PaaS* (Platform as a Service) is to develop and implement software applications created by a user via the Internet; *IaaS* (Infrastructure as a Service) is to provide on-line infrastructure where a customer may develop whatever software applications [19].

P.Mell and T.Grance define various service models of the cloud-based architecture (Fig.1) [16]. These models may be purposefully used for providing software access in educational institutions.

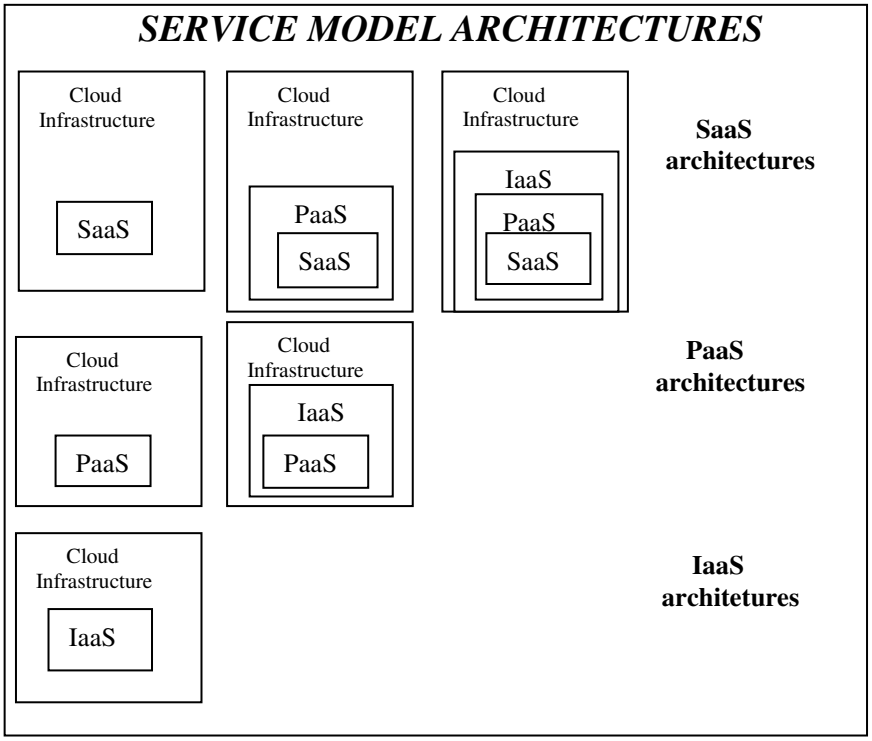


Fig. 1. Service model architectures (After P.Mell, T.Grance [16]).

There are also four *service deployment models* for cloud computing application that reflect the mode of the cloud infrastructure set up in a particular organization: *the corporate cloud* is owned or leased by the organization; *the cloud community* is a shared infrastructure used by a community; *the public cloud* is a mega-scale infrastructure that may be used by any person under some payment terms; *the hybrid cloud* is a composition of one or more models [4, 16].

In view of the different models for cloud service architecture, when choosing the most appropriate solution that is suitable for a particular organisation, both collective and individual users should be considered. Selecting the SaaS model in this respect can be justified by the fact that these services are the most accessible, although a thorough market analysis and educationally prudent choice of the necessary application that could fit learning or scientific purposes is to be made. This kind of service may be purposefully used by an individual and also a collective user.

At the same time, for the settlement of the ICT infrastructure of the institution by the PaaS or IaaS model, the selection and approval of the relevant cloud platform is necessary. This solution is concerned with a number of organisational issues, such as the formation of a special unit of ICT personnel skilled in setting up and deploying this infrastructure, configuring the hardware and software complexes, planning and working out the environmental design tasks, testing and approving its modules and components, filling it with the necessary resources, monitoring its implementation,

maintaining quality control, training the teaching staff, etc. [4]. In this case, given the results of recent research and the current trends in IT sector development, the use of hybrid service models appears to be a promising and prospective solution. The hybrid solutions are reported to be well-embedded into existing settlements provided by leading cloud suppliers, and this tendency is growing [9]. The hybrid cloud incorporates public and corporate cloud tools that do not necessarily exclude the involvement of software-as-a-service applications [19].

As shown in Fig.1, there are three approaches to implementation of learning software access in the SaaS architecture. In the first case (directly SaaS) the cloud platform deployment is not necessary in the educational institution this work is undertaken by a service provider. In both other cases the corporate or hybrid cloud deployment is needed. In this case the appropriate cloud platform (eg, Amazon Web Services, Microsoft Azure, Eucaliptus, Xen, WMWare etc.) is used to deploy the certain service model. In the process of cloud infrastructure configuration the guidelines are usually supplied by the vendor [1]. These guidelines contain a number of basic deployment scenarios that can be implemented. It is possible to build the cloud by means of different software and services but the basic notions are to be considered. One of the basic concepts of the cloud based learning environment configuration is the concept of a *corporate cloud* or a *virtual private cloud* (VPC). Sometimes the term is used not very clearly so as to describe the corporate cloud that may include a public and also a private part so being the hybrid one. Depending on the scenario chosen the certain model of software access is considered.

As a rule the cloud provider may propose services of several types. For example, it is possible to rent additional disk space (S3); the virtual machine (EC2), with certain parameters of the processor, memory, and disk capacity, it may be with some installed operational system and software; remote database (SimpleBD, RDS) and others [1]. Depending on the chosen scenario these resources are configured within the cloud infrastructure.

There are four types of scenario for the cloud infrastructure configuration that are mostly proposed by the provider [1]:

Scenario 1: VPC with only public subnet. The configuration of the virtual cloud under this scenario contains a single public subnet and Internet gateway so as to enable communication over the Internet. This configuration is recommended if it is necessary to run the single level, public web applications such as blogs, web sites [1].

Scenario 2: VPC with public and corporate subnets. The configuration for this scenario includes a public and corporate (private) subnet. This configuration is recommended if it is necessary to run a public web application, while internal servers are not publicly available. An example is a multi-website with the web servers situated in a public subnet, and the database servers to be in a corporate subnet. It is possible to configure the security services and routing so that the web servers could interact with the database servers [1].

Scenario 3: VPC with the public and corporate subnet components and virtual private network (VPN) access. The configuration for this scenario includes the virtual hybrid cloud with the public and corporate subnets and the virtual corporate gateway namely the VPN connections. In an educational institution may be own subnet, which

should be expanded by augmented cloud services, such as additional disk space, databases, virtual machines, network gateways, additional "desktops" and so on. VPN-connection is used to enable communication with this subnet. You can also create the virtual cloud subsystem (subnet virtual machine) with access to the corporate subnet via the Internet [1]. For this scenario, the multi-level applications with scalable web services may be run, some parts of these applications are in the public subnet, and another parts are in the corporate subnet, which is connected to own subnet through the VPN channel [1]. This allows you to keep some data in the limited access.

Scenario 4: VPC subnet with the corporate VPN access components. The configuration for this scenario includes the virtual corporate subnet and the virtual gateway to allow communication with own subnet through the VPN channel. This scenario is recommended if there is a need to expand own subnet into the cloud, as well as to provide direct access to the Internet from this subnet without making it "visible" from the Internet [1].

On the stage of environment design all the possible configurations were considered and the model of the Scenario 3 was chosen so as to provide the hybrid infrastructure were the corporate and public components were used (Fig. 2).

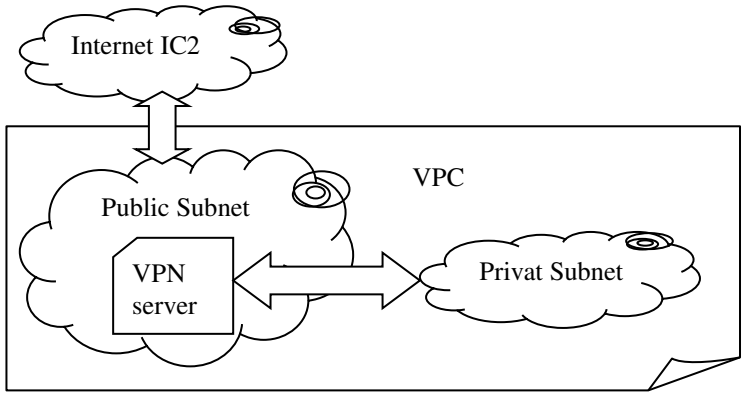


Fig. 2. The Hybrid Cloud configuration.

6 The Hybrid Service Model of Learning Software Access

To research the hybrid service model of learning software access, a joint investigation was undertaken in 2013–2014 at the Institute of Information Technologies and Learning Tools of the NAPS of Ukraine and Drohobych State Pedagogical University named after I.Franko. At the pedagogical university, the experimental base was established where the cloud version of the Maxima system (which is mathematical software), installed on a virtual server running Ubuntu 10.04 (Lucid Lynks), was implemented. In the repository of this operational system is a version of Maxima based on the editor Emacs, which was installed on a student's virtual desktop [21]. In this case, the implementation of software access due to the hybrid cloud deployment in Scenario 3 was organised.

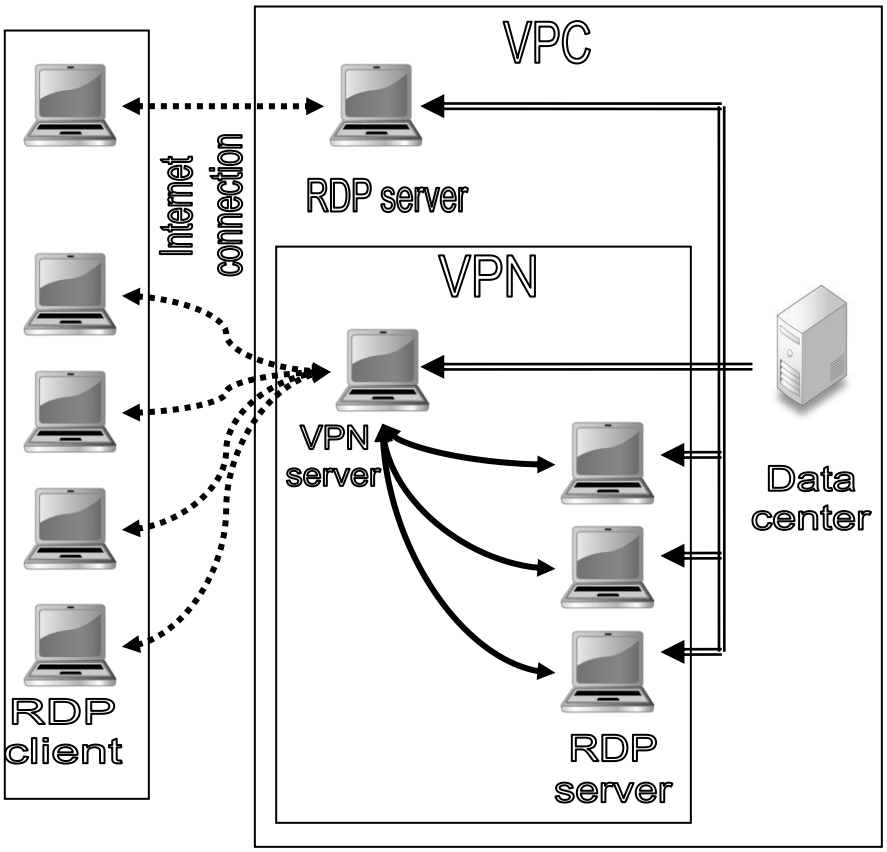


Fig. 3. The hybrid service model of the learning resources access.

In Fig.3, the configuration of the virtual hybrid cloud used in the pedagogical experiment is shown. The model contains a virtual corporate (private) subnet and a public subnet. The public subnet can be accessed by a user through the remote desktop protocol (RDP). In this case, a user (student) refers to certain electronic resources and a computing capacity set on a virtual machine of the cloud server from any device, anywhere and at any time, using the Internet connection.

In this case, a user's computer is the RDP-client, while the virtual machine in the cloud is the RDP-server. In the case of a corporate (private) subnet, a user cannot apply to the RDP-server via desktop because it is not connected to the Internet directly. Computers in the corporate subnet have Internet access via the VPN-connection, i.e. the gateway. Thus, these computers cannot be accessed from any

device, but only from the specially configured one (for example, a computer in the educational institution or any other device where the VPN-connection is set up) (Fig.3).

The advantage of the proposed model is that, in a learning process, it is necessary to use both corporate and public learning resources for special purposes. In particular, the corporate cloud contains limited access software; this may be due to the copyright being owned by an author, or the use of licensed software products, personal data and other information of corporate use. In addition, there is a considerable saving of computational resources, as the software used in the distributed mode does not require direct Internet access for each student. At the same time, there is a possibility of placing some public resources on a virtual server so the learner can access them via the Internet and use the server with the powerful processing capabilities in any place and at any time. These resources are in the public cloud and can be supplied as needed.

7 Implementation and Empirical Evaluation

In the joint research experiment held at Drohobych State Pedagogical University named after I.Franko, 240 students participated. The aim was to test the specially designed learning environment for training the operations research skills on the basis of the Maxima system. During the study, the formation of students' professional competence by means of a special training method was examined. The experiment confirmed the rise of the student competence, which was shown using the χ^2–Pearson criterion [21]. This result was achieved through a deepening of the research component of training. The experiment was designed using a local version of the Maxima system installed on a student's desktop.

The special aspect of the study was the expansion of these results using the cloud version of the Maxima system that was posted on a virtual desktop. In the first case study (with the local version), this tool was applied only in special training situations. In the second case study (the cloud version), the students' research activity with the system extended beyond the classroom time. This, in turn, was used to improve the learning outcomes.

The cloud-based electronic learning resource used in the experiment has undergone a quality estimation. The method of quality estimation in the joint laboratory of educational quality management with the use of ICT was used for this study [29]. The 25 experts were specially selected as having experience in teaching professional disciplines focused on the use of ICT and being involved in the evaluation process. The experts evaluated the electronic resource with such parameters as "Ease of access", "Ease of use" and "Usefulness". These parameters were chosen as they contribute to the accessibility of the cloud resource and the cloud-based learning in order to determine its feasibility and availability.

The problem was: is it reasonable and feasible to arrange the environment in a proposed way? There were three questions part of the access realisation mode (Table 1):

Table 1. The questionnaire.

	Parameter		Value
1	Ease of access	Is the electronic resource access easy and convenient?	0 (no), 1 (low), 2 (good), 3 (excellent)
2	Ease of use	Is the user interface clear and convenient?	0 (no), 1 (low), 2 (good), 3 (excellent)
3	Usefulness	Is this resource useful?	0 (no), 1 (low), 2 (good), 3 (excellent)

A four-point scale (0 (no), 1 (low), 2 (good), 3 (excellent)) was used for the questions. The 25 experts estimated two parameters, "Ease of access" and "Ease of use", and were invited to examine the resource. Experience using this resource in the learning process was not mandatory. The third parameter, "Usefulness", was estimated only by the seven experts who used the resource in the learning process. The results of the evaluation are shown in Fig.4.

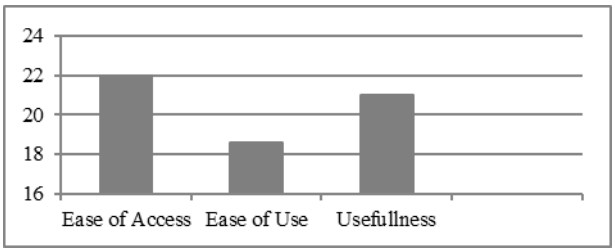


Fig. 4. The results of the cloud-based learning resource quality parameters evaluation.

The resulting average value was calculated for every parameter: "Ease of access" = 2.2, "Ease of use" = 1.86 and "Usefulness = 2.1. All criteria were weighted as one, and the total value was 2.1. This characterises the resource accessibility as sufficient for further implementation and use.

The advantage of the approach is the possibility to compare the different ways to implement resources with regard to the learning infrastructure. Future research in this area should consider different types of resources and environments.

8 Conclusion

The introduction of innovative technological solutions into the learning environment of educational institutions contributes to unified learning infrastructure formation and the growth of access to the best examples of electronic resources and services. ICT use is promising regarding learning settings that can advance and develop the tendencies of CC progress. For example, using the cloud-based models of environment design, virtualising applications, unifying infrastructure, integrating services, increasing the use of electronic resources, expanding collaborative forms of work, widening the use of the hybrid models of ICT delivery and increasing the quality of electronic resources.

References

1. Amazon Virtual Private Cloud. User Guide, API Version 2013-07-15, (2013)
2. Bard, G.V.: Sage for Undergraduates. AMS, (2015)
3. Bykov, V.: Models of Organizational Systems of Open Education. Atika, Kyiv (2009) (in Ukrainian)
4. Bykov, V.: Cloud Computing Technologies, ICT Outsourcing, and New Functions of ICT Departments of Educational and Research Institutions. Information Technologies in Education, 10, 8–23, (2011) (in Ukrainian)
5. Bykov, V., Lapinskii V.: The Methodological basis for creating and implementation of the electronic learning tools. Computer in school and in family, 2(98), 3-6, 2012.
6. Buyyaa, R., Chee Shin Yeoa, Venugopala, S., Broberga, J., Brandicc, I.: Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility. Future Generation Computer Systems, 25(6), 599–616, (2009)
7. Cusumano M.: Cloud computing and SaaS as new computing platforms. Communications of the ACM, 53(4), 27-29 (2010)
8. Doelitzscher, F., Sulistio, A., Reich, Ch., Kuijs, H., Wolf, D.: Private cloud for collaboration and e-Learning services: from IaaS to SaaS. Computing, 91, 23–42, (2011)
9. The Future of Cloud Computing: 4th Annual Survey 2014. The North Bridge Future Of Cloud Computing Survey In Partnership With Gigaom Research, http://bit.ly/2014FutureCloud (2014)
10. Gold, N., Mohan, A., Knight, C., Munro, M.: Understanding service-oriented software, Software, IEEE, 21(2), 71 – 77, (2004)
11. James, M.: Free Sage Math Cloud - Python and Symbolic Math. the I Programmer, Friday, http://i-programmer.info/news/202-number-crunching/6805-free-sage-math-cloud-python-and-symbolic-math-.html (2014)
12. Hashmi1 S.I., Clerc V., Razavian M., and others: Using the Cloud to Facilitate Global Software Development Challenges. 2011 Sixth IEEE International Conference on Global Software Engineering Workshops, (2011)
13. Lakshminarayanan, R., Kumar, B., Raju, M.: Cloud Computing Benefits for Educational Institutions. In Second International Conference of the Omani Society for Educational Technology, Muscat, Oman: Cornell University Library, http://arxiv.org/ftp/arxiv/papers/1305/1305.2616.pdf (2013)

14. Maamar, Z., et al.: An approach to engineer communities of web services: Concepts, architecture, operation, and deployment. International Journal of E-Business Research (IJEBR), 5(4), 1-21, (2009)
15. Matheson C., Matheson D.: Access and Accessibility in E-Learning. Applied E-Learning and E-Teaching in Higher Education. Ed. by Donnelly R., McSweeney F., Hershey New York, 130-151, (2009)
16. Mell, P., Grance T.: Effectively and Securely Using the Cloud Computing Paradigm. NIST, Information Technology Laboratory, 10-7-2009, (2009)
17. Qing Li, Ze-yuan W., Wei-hua Li, Jun Li, Cheng Wang, Rui-yang Du.: Applications integration in a hybrid cloud computing environment: modelling and platform. Enterprise Information Systems, 7(3), 237-271, (2013)
18. Smith, A., Bhogal J., Mak Sharma: Cloud computing: adoption considerations for business and education. 2014 International Conference on Future Internet of Things and Cloud (FiCloud), (2014)
19. Shyshkina, M.: Innovative Technologies for Development of Learning Research Space of Educational Institution. Information Technologies and Society, 1, http://ifets.ieee.org/russian/depository/v16_i1/pdf/15.pdf (2013) (In Russian)
20. Shyshkina, M.: Emerging Technologies for Training of ICT-Skilled Educational Personnel. Communications in Computer and Information Science, Berlin-Heidelberg, Springer-Verlag, 412, 274-284, (2013)
21. Shyshkina M. U. P. Kohut, I. A. Bezverbnyy. Formation of professional competence of computer science bachelors in the cloud based environment of the pedagogical university. Problems of modern teacher preparation, Uman, FOT Zhovtyy O.O., 9, part 2, 136-146 (2014) (in Ukrainian)
22. Sultan, N.: Cloud Computing for Education: A New Dawn? Int. J. of Information Management, 30, 109–116, (2010)
23. Tuncay, E.: Effective use of cloud computing in educational institutions. Procedia - Social and Behavioral Sciences, 2(2), 938–942, (2010)
24. Turner, M., Budgen, D., Brereton, P.: Turning software into a service. Computer, 36 (10), 38-44, (2003)
25. Vaquero L. M.: EduCloud: PaaS versus IaaS cloud usage for a n advanced computer science course, IEEE Transactions on Education, 54(4), 590-598, (2011)
26. Vouk, M.A., Rindos, A., Averitt, S.F., Bass, J. and others: Using VCL technology to implement distributed reconfigurable data centers and computational services for educational institutions. VCL/Reconfigurable. Data Centers & Clouds/NCSU/V19-Draft Feb-2009 1-27, (2009)
27. Wick D.: Free and open-source software applications for mathematics and education. Proceedings of the twenty-first annual international conference on technology in collegiate mathematics, 300-304, (2009)
28. Zhang, J., and others: A Framework of User-Driven Data Analytics in the Cloud for Course Management. Proceedings of the 18th International Conference on Computers in Education, S. L. Wong et al., Eds., Putrajaya, Malaysia, Asia-Pacific Society for Computers in Education, 698-702, (2010)
29. Zaporozhchenko, Yu., Shyshkina M., Kravtsov G.: Prospects of the development of the modern educational institutions' learning and research environment: to the 15th anniversary of the Institute of information technologies and learning tools of NAPS of Ukraine. Informational Technologies in Education, 19, 62-70, (2014).

Methods and Technologies for the Quality Monitoring of Electronic Educational Resources

Hennadiy Kravtsov[1]

[1] Kherson State University
27, 40 rokiv Zhovtnya St., 73000, Kherson, Ukraine
kgm@ksu.ks.ua

Abstract. Support of the quality of training is one of the main objectives of the university system. The results of modeling the quality management system of electronic educational resources (EER) on the basis of the analysis of its elements are presented. The subject of the study is the EER quality monitoring. Technologies for EER quality monitoring are based on the method of expert evaluations. The criterion of EER quality is considered as the weighted average value of quality indicators. The weights of EER types and indicators of EER quality for their types are evaluated in pedagogical experiment. Results of experiment confirmed the assumption that the method of expert evaluations can be the basis for the EER quality monitoring. Concordance method is used to assess the degree of consensus of experts on the factors: weights of EER types, parameterization of EER quality indicators, and weighted average criterion of EER quality. The model of quality management system is shown in the example of assessing the quality of the distance learning system resources.

Keywords. quality management system, electronic educational resources, monitoring of quality, distance learning system «Kherson virtual university».

Key Terms. QualityAssuranceMethodology, StandardizationProcess, KnowledgeManagementMethodology, KnowledgeManagementProcess, TeachingMethodology.

1 Introduction

Electronic educational resources (EER) is object of quality management system of the educational process with the use of ICT [1, 2]. There are two main approaches to the concept of quality EER: compliance with standards and customer requirements. Therefore it is necessary to take into account two aspects: compliance with educational standards and meeting the requirements of students and teachers of the university. The compatibility with international standards IMS, SCORM can be chosen as a criterion for EER quality.

Improving the EER quality is the main purpose of the quality management system (QMS) [3]. Implementation of QMS in institutions can improve processes by establishing the effective and efficient management systems. Thus, EER quality management

provides tools, methods and technologies for the continuous improvement of the educational process. This improves performance, reduces the costs and ultimately increases the competitive advantages of the institution.

Standards ISO 9000/9001 and ISO 29990 represent one of the models of management of the institution to ensure the quality of the educational process [4]. Monitoring is an essential tool of evaluating the quality of the educational process, in particular the quality of the EER. The EER quality monitoring is understand as continuous process of observation and recording EER parameters and their subsequent evaluation. Quality monitoring provides expert advice according to the estimating procedure of the EER.

Because EER are classified as electronic educational editions and at the same time they are software products then EER quality monitoring should be multilevel taking into account their classifications.

The basic types of electronic educational resources for EER quality monitoring should be assigned. For each EER type the weight factors and quality indicators should be offered. The general criterion of quality electronic resources should be used to assess their quality. It is average weighted characteristic of quality and takes into account the weights of resource types and their relative quality indicators. The assessment of EER quality monitoring is given by a corresponding university commission of experts [1].

Task of the present work is the analysis, calculation and optimization of parameters of EER quality management system with use of methods for the analysis of complex systems [5].

2 Model of EER Quality Management System

The EER quality management system is a structural element of architecture of education quality management system in the higher educational institution. It plays a feedback role in EER quality management system of educational process.

The structure of EER quality management system is presented on figure 1 [1].

Let's list the basic elements of quality management system of electronic resources of learning.

Assessment of EER quality underlies a quality management system of electronic resources of learning. For an assessment of EER quality it is necessary:

- to carry out monitoring for control of EER quality on a fixed basis;
- to have a feedback with users of EER for the account of wishes in their improvement from positions methodical and program-technology requirements.

It is necessary to develop these criteria of EER quality for carrying out of monitoring of quality. The university council of experts confirms the criteria of EER quality developed by the methodical commissions. The university council of experts also confirms the recommendations about improvement qualities of EER received as a result of the analysis of users' responses in Feedback system.

Results of an assessment of EER quality should be used on the one hand for improvement of their substantial part and satisfaction to technology requirements, and on the other hand for publication of a rating of electronic learning resources that also promotes the increase of their quality.

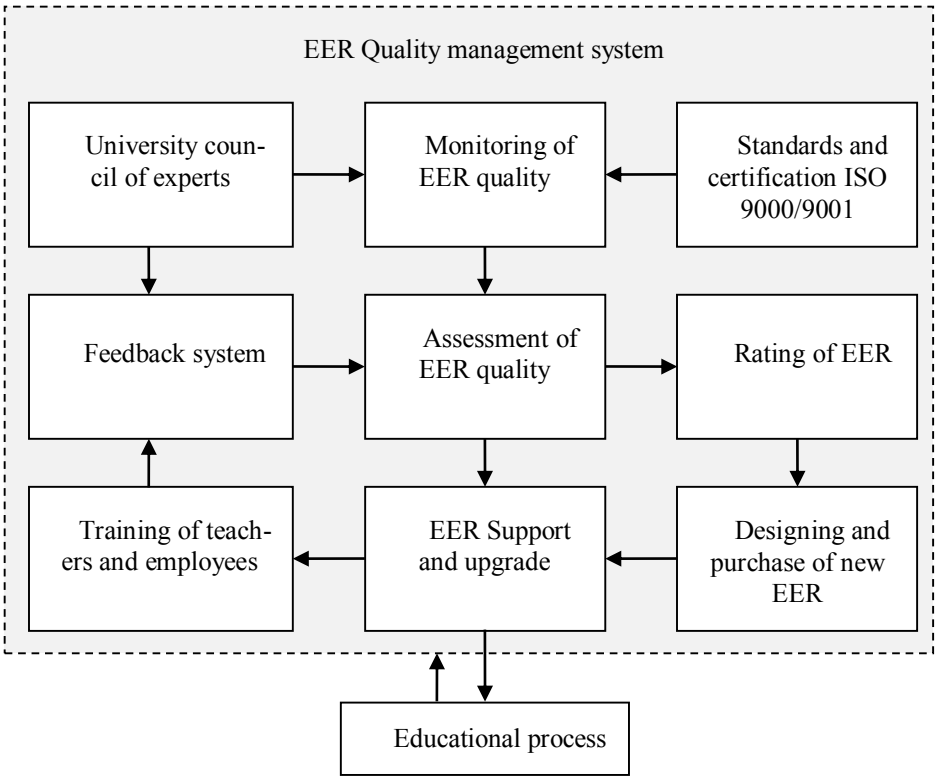


Fig. 1. Structure of EER quality management system.

Monitoring of EER quality has a leading role at their assessment of quality. The analysis of electronic resources of learning shows, that they have the following classification: to a functional character they can be referred to learning editions, under the form of representation they belong to a category of electronic editions, on the technology of creation they represent software product [5]. Therefore the monitoring of quality of electronic educational resources should be multi-criterion and multilevel according to their classification. The satisfaction requirement to the universal international standards that are IMS, SCORM [6] is the uniting attribute of multilevel monitoring of EER quality.

At monitoring of EER quality it is necessary to consider, the certain typological model of system of educational editions for high schools which includes four groups of the educational information resources differentiated to a functional sign, defining their value and a place in educational process has affirmed [7]: learning-methodical, training, auxiliary and supervising.

At monitoring of EER quality by criterion of compatibility with educational standards at definition of quality indicators it is possible to use specifications IMS which describe information model of educational objects. These specifications define the

standardized set of information blocks which contains data about an educational resource. The IMS-package which contains educational object consists of two main elements [6]:

- the IMS-manifesto – a special file which describes the base resources, the maintenance and the organization of educational object (it is represented in language XML);
- the physical files which make educational object.

At monitoring of EER quality it is necessary to consider their typical classification: electronic textbooks and methodical manuals, practical and virtual laboratory works, tests and training simulators, etc.

Among all EER the special role is played by a distance learning course. It is the basic educational object which is used in distance learning. It is compound training object which unites various EER for the purpose of the organization of learning process with use of special program environments – Distance Learning System (DLS). The example of such program environment which allows to create, keep and use distance courses, is DLS «Kherson Virtual University» [7].

The criterion of EER quality is considered as the average factor of quality $K = (\alpha_1 k_1 + \alpha_2 k_2 + ... + \alpha_n k_n)/n$, where α_i – average value of quality indicators, k_i – value of weight factor of i-type resource.

The general relative average criterion of EER quality can be calculated under the formula [1]

$$K = \sum_{i=1}^{N} a_i t_i \,. \tag{1}$$

Here $a_i = n_i \cdot \gamma_i$ – the quality metrics, $\gamma_i = \sum_{j=1}^{m_i} k_{ij} \,/k_{iM}$ – average factor of quality, n_i – weight factor, m_i – quantity of metric indicators of quality, k_{ij} – j-indicator of quality, k_{iM} – the maximum value of an indicator of quality, t_i – the generalized factor of quality of i-type resource, N – quantity of EER.

The Feedback system serves as the tool for the organization of flexible and all-round polls of opinions of students and teachers of university. Usually the system takes questioning in an automatic mode. The generalized assessment of EER quality was received after statistical processing of results of questioning of users, it gives the opportunity to consider the degree of their demand at quality monitoring.

Standards and certification ISO 9000/9001. Certification is a documentary acknowledgement of conformity of production to certain requirements, concrete standards or specifications. It is necessary to notice, that conformity to standard ISO 9000/9001 does not guarantee high EER quality. However conformity to requirements and recommendations of these standards is a necessary condition of high quality of resources of training. The certificate of conformity ISO 9001 is acknowledgement of satisfaction to standard requirements.

Standard ISO 9000/9001 is fundamental, the terms and definitions accepted in it are used in all standards of a series 9000. This standard is a basis for understanding of base elements of QMS according to ISO standards.

Requirements of standard ISO 9000/9001 can be used as criteria at the organization

and carrying out of monitoring of EER quality.

University council of experts. In the control system of EER quality the university advisory council is the body which is responsible for adequacy assessment of EER quality taking into account all criteria and indicators of quality. It adopts the Regulation about ERR quality management system, defines the criteria of their quality, forms rules of carrying out and confirms results of an assessment of quality, and also plans actions for improvement of EER quality.

The university advisory council defines the procedure of carrying out of monitoring of EER quality. It confirms the list of criteria of quality, their weight factors and values of indicators of quality according to (1).

Support and upgrade of EER is the important part of work in QMS for improvement and optimization of EER software at its use in educational process. Support EER is one of the phases of the software lifecycle. The software logs the detection correction, and add new functionality to increase efficiency. Support software is defined by standard IEEE Standard for Software Maintenance (IEEE 1219), and the life cycle standard is specified ISO 12207.

The important factor of increase of efficiency usage of EER is training of users and maintenance them with regular support at work with the current software version.

3 Integrated and differentiated approaches in modeling and use of EER quality management system

The control system of EER quality is a model which describes the business process including actions and activity of services of university according to functionality of structure described above the scheme of EER quality management (fig. 2). It is necessary to notice, that some elements of this system possess the property of close interrelation and have various degrees of influence on it. Thus some elements of the system (for example, «University Advisory Council» and «Standards and Certification ISO 9000/9001» at monitoring of EER quality) can be united in groups which we will name services. Therefore for the purpose of allocation of major factors of a quality control system, influencing quality of its work, on the basis of its structure (fig. 2) we form three main places of maintenance of EER quality: service of quality monitoring, service of quality assessment and EER support and upgrade service. We will define structure, primary goals, requirements and expected results of work of these services.

The Service of quality monitoring is intended for the organization and carrying out of EER quality monitoring which are used in educational process, by criterion of their conformity to the international educational standards. The University advisory council defines the order and rules of carrying out of monitoring of EER quality.

Service tasks: the coordination of parameters and development of criteria of EER quality, taking into account the requirements of standards, carrying out of analysis of EER by the developed and coordinated criteria.

Requirements: carrying out monitoring on fixed basis, completeness of coverage of all kinds of EER, objectivity of application of criteria of quality.

Expected results: data of the analysis of EER characteristics for their assessment of

quality.

The Service of an assessment of quality makes EER assessment on the basis of the confirmed criteria taking into account the opinion of users – both students, and teachers. Feedback system can be used for automation of carrying out polls and processing of results.

Service tasks: to assess of EER quality by the developed and coordinated criteria on the basis of the analysis of their characteristics for maintenance of formation of rating.

Requirements: objectivity, publicity, competitive character.

Expected results: on the basis of quality assessment to generate the list of reclamations to electronic resources of learning for performance of works on their elimination and to make rating of EER for increasing of motivation of authors of resources for improvement of their quality.

The Service of EER support and upgrade carries out the organization, planning and performance of works on improvement of their quality by correction of the noticed lacks, realization of new didactic properties and possibilities of electronic resources of learning. Experts of this service give consulting services in acquiring new EER, and also take part in training of teachers and employees to use them.

Service tasks: on a constant basis taking into account an assessment of EER quality to perform works on their upgrade and as much as possible to satisfy inquiries of users.

Requirements: operatively, qualitatively and full performance of works.

Expected results: upgrade and introduction new and improved EER in educational process of university.

3.1 Analysis EER QMS by criteria of its elements importance

Services of control system of EER quality provide the consecutive process of their monitoring, assessment of quality and support. Thus Feedback system plays a feedback role in this process. On fig. 2 the function chart of work of services of EER QMS is presented.

According to methods of the theory of automatic control we will designate through $W_i(p)$ - transfer functions of EER quality of corresponding services ($i = 1,2,3$) and Feedback system ($i = 4$) [8]. According to rules of calculation of consecutive connection of links of system and taking into account Feedback system transfer function of opened system $W(p)$ is expressed through the transfer functions of corresponding links $W_i(p)$ under the formula

$$W(p) = \frac{W_1(p) \cdot W_2(p) \cdot W_3(p)}{1 \pm W_2(p) \cdot W_3(p) \cdot W_4(p)}. \qquad (2)$$

It is necessary to notice, that the Feedback system can play a role both local negative (–), and local positive (+) feedback. Thus the role of a negative feedback is more significant and more often is used in work of EER QMS as the main mission of EER QMS consists in revealing of resources of poor quality and their upgrade. At the same time the system can be in a status of action of a local positive feedback in case of a mode of

popularization of the best practices on creation qualitative EER.

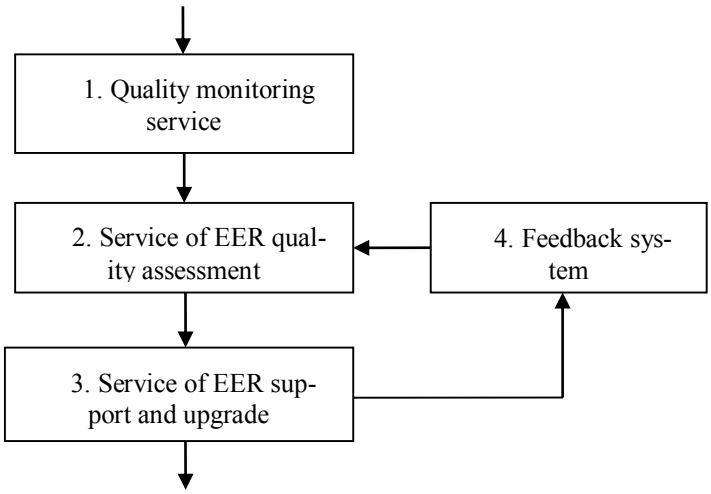


Fig. 2. The scheme of service functionality in the EER quality management system.

With sufficient degree of generality it is possible to consider the model of ideal strengthening of links of system. Then $Wi(\mathrm{p}) = k_i$ (i = 1,2,3,4), where k_i-factors of improvement of EER quality of corresponding i-links of system. Generally for factor k improvement of EER quality of all the QMS from (2) we have expression

$$k = \frac{k_1 \cdot k_2 \cdot k_3}{1 \pm k_2 \cdot k_3 \cdot k_4}. \tag{3}$$

Considering, that the control system of EER quality is a global feedback in architecture of control system of learning quality, the condition performance suffices for maintenance of improvement of electronic resources quality $k > 1$ or

$$k_1 \cdot k_2 \cdot k_3 > 1 \pm k_2 \cdot k_3 \cdot k_4. \tag{4}$$

The correlation (3) together with a condition (4) allows to apply the differentiated approach to the account of degree of importance of elements of EER QMS, and also to optimize parameters of this system.

3.2 Methods of calculation and optimization of parameters of EER QMS

For the purpose of optimization of parameters of EER QMS we will apply the method of consecutive allocation of the major elements of system by criterion of their influence on system from the point of view of EER quality. In considered above the model of ideal strengthening of links of system the factors of improvement of EER

quality can act as weight factors of the importance of elements of EER QMS of learning. The optimum combination of values of these factors will promote the optimization of operating modes of all control system by quality of electronic resources. In practice factors k_1, k_2, k_3 and k_4 are not the determined parameters, and have properties of random variables with the known law of distribution therefore at modeling of optimum statuses of EER QMS it is necessary to apply statistical methods of calculation and optimization of parameters of system.

As example of use of statistical methods of calculation and optimization of parameters of system the calculation of an average of distribution of factor k improvement of EER quality depending on average of distributions of factors k_i can serve. Optimization of dispersion of values k is realized by imposing of restrictions on known values of average of distributions and mean square deviations of factors k_i.

4 Implementation and Empirical Evaluation of EER quality management system

4.1 Method of expert evaluations of EER quality

In assessing the EER quality by the form of organization the method of collective estimation is used with collective expert opinion. This method is used to obtain quantitative estimates of the quality characteristics, parameters and properties. Analysis of expert assessments involves filling each individual expert appropriate form, the results of which are a comprehensive analysis of the problem situation and possible solutions. The results of peer reviews are issued as a separate document.

The purpose of peer reviews of EER quality is an evaluation of EER quality indicators with international, national and industry standards, the EER quality monitoring, quality of the learning process through the use of qualitative EER and processing methods, criteria and forms for certification e-learning.

Objects and *parameters* of EER assessment:

- Classification of EER types.
- The weight factors of EER types (EER relative priority for their type).
- Factors and criteria of EER quality for their types.

The following forms of expertise processing of EER quality are:

1. Definition of the competence of experts and the formation of the expert committee.
2. Evaluation of weight factors ranging of EER types.
3. Parameterization of EER quality indicators.
4. Expertise processing of EER quality.
5. Study the adequacy of the results of expertise.

Expert committee is created for the EER expertise with use of peer reviews method. *Delphi method* is used in the formation of the expert committee and expertise processing [9]. Top teachers, methodologists and researchers of higher education institutions are involved in the commission of experts.

Since EER are classified as electronic publications for educational purposes and they

are software products, the examination of the quality of electronic educational resources should be layered with regard to their classifications. Therefore, the EER quality should be analyzed by the software and technological, psychological, pedagogical and ergonomic features.

EER quality indicators are derivative of the requirements for them. Meeting the requirements of program-technological, psychological, pedagogical and ergonomic ones are a measure of EER quality assessment in determining their quality indicators

In this case the development of tools is based on modern fulfilled hygiene, ergonomic and technical and technological standards to the use of computer technology and is governed by existing regulations or standards. You can ask to have developed technology expertise of EER quality indicators that can be fully regulated in detail. However, there are problems of evaluating these indicators related to obsolescence of existing standards and the fact that definition of quality are not further developed.

4.2 The EER quality monitoring in educational institutions

Monitoring and evaluation (M&E) is a process that helps improving performance and achieving results. Its goal is to improve current and future management of outputs, outcomes and impact [10]. Consider the EER quality monitoring by the example of DLS «Kherson Virtual University» [7].

Formation of the commission of experts. Determining the validity of each of the three subjects of the educational process was made by expert evaluation method. 25 qualified experts (university teachers, graduate students, methodologists) was joined the independent expert committee.

To define a point of evaluation for each subject Delphi method (for members of the expert committee conditions for an independent individual work were created) was used. The statistical processing of the results, which were presented to experts for final approval, had been conducted.

Construction of weights ranging of EER types. The weight factor of EER type is a numerical coefficient, a parameter that determines the value, the relative importance of this EER type than other types that are classified EER on functional grounds.

Table 1 shows an example of a possible evaluation of EER weighting coefficients values according to their types.

Table 1. The weighting factors of EER types.

#	Name of EER Type	Description	Weighting factor
1	Electronic textbooks and books	Full course of lectures, encyclopedia	24,9
2	Lectures notes, laboratory and practical work notes	Lectures annotations, laboratory and practical work annotations	21,2
3	Lecture Presentation	Author lecture in Power Point format	16,0
4	Video Lecture	Author lecture in video format	19,5

5	Audio Resource	Author EER in audio format	15,1
6	Learner's guide	Electronic learner's guide in discipline	26,9
7	Guidance for conducting seminars and laboratory works	Full description of seminars, laboratory and practical works	18,8
8	Laboratory work	Virtual laboratory works in discipline	21,3
9	Test	Full set of questions with indicating correct answers	17,6
10	Library of electronic visual aids	The library of visual learning objects in a graphical format	26,3
11	Collection of tasks, exercises, vocabulary	Author's electronic resource	25,9
12	Training computer game	Author's electronic resource	23,9
13	The work program of the course	Approved author's work program in discipline	19,6
14	Questions to exam/credit, self-control	In accordance with the work program	17,2
15	Print and Internet resources	Basic and advanced print and online resources of discipline with active hyperlinks	18,4
16	Distance course in the discipline	Correspond to international standards	98,1

Parameterization of EER quality indicators

The EER quality indicator is a numerical parameter that determines the evaluation the EER under its qualitative characteristic (can be used a five point Likert's system). Also the EER types are specified, which are measured by this indicator. Filling out the list of EER quality indicators and their attachment to the EER types is held after approving the list of EER types.

Parameterization of EER quality indicators means evaluation of quality by scaling method [11]. Table 2 shows an example of evaluation of EER quality indicators under their quality point scale.

Table 2. The EER quality indicators.

Name of EER quality indicator. Description	What EER types is applied to	Quality characteristics	Quality indicator
Completeness of methodical support of discipline	All types	1. Full	5
		2. Incomplete	4
		3. Average	3

		4. Below Average	2
		5. Inadequate	1
Authorship of EER	All types	1. Full	5
		2. collaboration	3
		3. Plagiarism	0
EER compliance with state education standards	All types	1. Full	5
		2. Incomplete	3
		3. No	1
EER compliance with international standards:IMS, SCORM, IEEE etc.	1, 6, 9, 16	1. Full	5
		2. Incomplete	3
		3. No	1
EER compliance to work program content	All types	1. Full	5
		2. Incomplete	3
		3. No	1
Completeness of presenting educational material	1, 2, 3, 6, 16	1. Full	5
		2. Short	4
		3. Note	3
		4. Plan	1
The use of resources with respect to the maximum possible	All types	1. High	5
		2. Mediate	3
		3. Low	1
Structuring and formatting of educational material	1, 2, 3, 6, 7, 8, 11, 16	1. Yes	5
		2. Partially	3
		3. No	1
Text ergonomics	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 16	1. Quality	5
		2. Mediate	3
		3. Poor	0
Hypertext links use	1, 2, 3, 6, 7, 10, 15, 16	1. Yes	5
		2. No	0
Use of visual methods in material	1, 2, 3, 6, 7, 8, 10, 12, 15, 16	1. Quality	5
		2. Mediate	3
		3. Poor	0
Using multimedia	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 15, 16	1. Quality	5
		2. Mediate	3
		3. Poor	0
The use of interactive systems and modules, simulation	1, 2, 3, 6, 7, 8, 9, 10, 11, 12, 15, 16	1. Yes	5
		2. No	0

Using testing, the ability to control knowledge, self-control	1, 2, 3, 6, 7, 8, 9, 11, 12, 15, 16	1. Yes 2. No	5 0
Use file formats standard	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16	1. Yes 2. Partially 3. No	5 3 0
Use tables, charts, figures	1, 2, 3, 6, 7, 8, 9, 10, 11, 15, 16	1. Yes 2. No	5 0
Compliance learning material to knowledge level of students	All types	1. Yes 2. No	5 0
Purpose of educational material to an appropriate audience	All types	1. Yes 2. No	5 0
Free access to educational material	All types	1. Yes 2. No	5 0
The stylistic correctness of teaching learning material	1, 2, 3, 6, 7, 8, 10, 11, 15, 16	1. Quality 2. Mediate 3. Poor	5 3 0
The sequence of teaching learning material	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 13, 14, 15, 16	1. Quality 2. Mediate 3. Poor	5 3 0
Validity of test, tutorial	1, 3, 6, 7, 8, 9, 16	1. Yes 2. No	5 0
Automatic processing of test results and knowledge control	1, 2, 3, 6, 7, 8, 9, 11, 12, 16	1. Yes 2. No	5 0
Accessibility of used informational resources	1, 2, 3, 6, 7, 8, 11, 15, 16	1. Yes 2. No	5 0
Matching of EER components to psychological requirements	All types	1. Quality 2. Mediate 3. Poor	5 3 0

The study of the adequacy of experiment results

Expert evaluation of the EER quality can be considered sufficiently reliable only when a good consistency of expert answers. Therefore, the statistical processing of the results of experts evaluations should include an analysis of consensus of experts. Concordance method is used to assess the degree of consensus of experts on the factors: weights of EER types, parameterization of EER quality indicators, and average factor of EER quality [12].

Experts were asked to complete the table 1 for peer review weighting factors of EER types. The values of the weighting factors were selected from 100 point scale. The results of the survey of experts are presented in Table 3.

Table 3. Expert data on weights of EER types.

Ex-pert	ERR Types															
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15	#16
1	2	11	14	7	9	4	5	3	6	12	8	10	13	15	16	1
2	2	11	14	4	8	9	10	5	6	16	3	7	12	13	15	1
3	3	10	12	7	8	2	6	4	5	11	9	13	14	15	16	1
4	4	7	10	6	8	2	5	3	9	12	11	14	13	15	16	1
5	2	10	14	9	8	3	12	4	5	6	7	11	13	15	16	1
6	3	9	10	8	7	2	6	4	5	11	12	14	15	13	16	1
7	2	12	11	8	10	5	4	3	7	13	6	9	14	15	16	1
8	3	8	13	4	7	2	6	9	5	12	10	11	16	14	15	1
9	4	10	11	6	8	2	3	5	7	12	9	13	14	15	16	1
10	2	5	13	8	7	3	6	4	10	11	9	12	16	14	15	1
11	2	11	12	6	10	5	4	3	8	13	7	9	14	15	16	1
12	2	9	11	7	8	3	4	5	6	10	12	13	14	15	16	1
13	3	10	9	8	13	2	5	4	6	12	7	11	14	16	15	1
14	3	12	13	7	9	2	4	5	6	11	8	10	14	16	15	1
15	2	8	12	6	10	3	5	4	7	13	9	11	14	15	16	1
16	5	10	11	8	6	2	3	9	4	12	7	13	14	15	16	1
17	2	9	10	7	8	4	14	3	6	12	5	11	13	16	15	1
18	2	13	11	8	10	5	4	3	7	14	6	9	12	15	16	1
19	2	6	13	11	9	3	4	5	8	12	7	10	14	15	16	1
20	4	11	10	7	8	2	3	5	6	13	9	12	15	16	14	1
21	2	12	7	8	13	3	4	5	6	11	10	9	14	15	16	1
22	5	14	13	9	2	3	6	4	7	11	8	12	16	10	15	1
23	3	11	14	7	9	5	4	2	6	10	8	13	12	16	15	1
24	2	12	13	8	7	3	6	4	5	11	9	10	16	14	15	1
Δ_i	-138	37	77	-30	-2	-125	-71	-99	-51	77	-8	63	132	149	169	-180

Concordance coefficient W is calculated according to the formula proposed by Kendall [12]

$$W = \frac{12S}{m^2(n^3 - n)}. \quad (5)$$

Here $S = \sum_{i=1}^{n} \Delta_i^2 = \sum_{i=1}^{n} \left\{ \sum_{j=1}^{m} x_{ij} - \frac{1}{2} m(n+1) \right\}^2$, m – number of experts, n – the number of objects of examination (e.g., EER types), x_{ij} – assessment of the i-object by j-expert. Coefficient of concordance may vary between 0 and 1. If $W = 1$, all experts gave the same evaluations for all objects, if $W = 0$, the evaluations of experts are not coordinated.

Using the formula (5) we calculated that coefficient $W = 0{,}872$ and it is significantly different from zero, so we can assume that among experts there is objective concordance. Given that the value of $m(n-1)W$ is distributed according to χ^2 with $(n-1)$ is the degree of freedom, then $\chi_W^2 = \frac{12S}{m \cdot n \cdot (n+1)} = 314{,}1$. Comparing this value with the tabulated value χ_T^2 for $n - 1 = 15$ degree of freedom and significance level $\alpha = 0{,}01$, we find $\chi_W^2 = 314{,}1 > \chi_T^2 = 30{,}578$. Therefore, the hypothesis of consistency of expert evaluations confirmed according to Pearson.

Thus, the results of pedagogical experiment confirmed the assumption that the method of expert evaluations can be the basis for the EER quality monitoring.

5 Conclusions and Outlook

The system of EER quality monitoring is based on the multi-criterion analysis of conformity of these resources to the educational standards. Criterion of EER quality compatibility with standards IMS, SCORM can be chosen.

Criteria of EER quality are described on a basis the multi-criterion analysis taking into account EER compatibility with the international standards.

The basic types of electronic resources of educational appointment for carrying out of monitoring of EER quality are allocated. For each type of EER their weight factors and quality indicators are offered. The criterion of quality of an electronic training resource which is the average characteristic of quality is developed.

Technologies for EER quality monitoring is based on the method of expert evaluations. The criterion of EER quality is considered as the weighted average value of quality indicators. The weights of EER types and indicators of EER quality for their types are evaluated in pedagogical experiment. Results of experiment confirmed the assumption that the method of expert evaluations can be the basis for the EER quality monitoring. Concordance method is used to assess the degree of consensus of experts on the factors: weights of EER types, parameterization of EER quality indicators, and weighted average criterion of EER quality. The model of quality management system is shown in the example of assessing the quality of the distance learning system resources.

The offered system of an assessment of ERR quality is not unique and supposes additions and updating. The assessment of monitoring of EER quality is given by a corresponding commission of experts of university.

The method of testing is used for the experimental verification of the results of expert evaluation of the ERR quality. Electronic educational resources are subject to testing by means of their actual use in the educational process. As a result of comprehensive testing, a system of adjustments is formed to improve the ERR. The process of testing and further development of electronic educational resources is an iterative cyclical process. It should continue until achieving compliance with the ERR quality requirements. Therefore, the process of testing is an element of the quality management system of electronic educational resources. That study of ERR quality management system with their testing in educational process is the prospect of further work.

References

1. Kravtsov H.M. Design and Implementation of a Quality Management System for Electronic Training Information Resources / In: Ermolayev, V. et al. (eds.) Proc. 7-th Int. Conf. ICTERI 2011, Kherson, Ukraine, May 4-7, 2011, CEUR-WS.org/Vol-716, ISSN 1613-0073, P.88-98, online CEUR-WS.org/Vol-716/ICTERI-2011-CEUR-WS-paper-6-p-88-98.pdf. – P. 88 – 98
2. H. Kravtsov. Structure of the Management System of Quality of Electronic Learning Resources / Information Technologies in Education. 10th Issue. – Kherson. – 2011.– P. 94-101
3. Peris-Ortiz, M., Álvarez-García, J., Rueda-Armengot, C.: Achieving Competitive Advantage through Quality Management. Springer International Publishing Switzerland (2015). – URL: http://www.springer.com/gp/book/9783319172507
4. ISO 9000 - Quality Management. – URL: http://www.iso.org/iso/home/standards/management-standards/iso_9000.htm, Learning services for non-formal education and training
5. Bykov V. Yu. Models of the Open Education Organizational Systems: Monograph. – Kyiv: Atika, 2009. – 684 p.: ill.
6. H. Kravtsov, D. Kravtsov. Knowledge Control Model of Distance Learning System on IMS Standard / Innovative Techniques in Instruction Technology, E-learning, E-assessment, and Education. – Springer Science + Business Media V.B. – 2008. – P.195 – 198.
7. H. Kravtsov. Evaluation Metrics of Electronic Learning Resources Quality / Information Technologies in Education. 3d Issue. – Kherson. – 2009. – P. 141 – 147.
8. Jay C. Hsu, Andrew U. Meyer. Modern Control Principles and Applications. McGraw-Hill (1968)
9. Rowe, G. & Wright, G. Expert Opinions in Forecasting: The Role of the Delphi Technique. In: J.S. Armstrong (Ed.), Principles of Forecasting - A Handbook for Researchers and Practitioners, pp. 125-144. Boston, MA; Kluwer Academic Publishers (2001)
10. Wikipedia – The Free Encyclopedia. – URL: https://en.wikipedia.org/wiki/Monitoring_and_Evaluation
11. Kolen, Michael J., Brennan, Robert L. Test Equating, Scaling, and Linking. Methods and Practices. Springer-Verlag New York (2004)
12. Kendall M. Rank Correlation Methods, Charles Griffen & Company, London (1948)

Realisation of "Black Boxes" Using Machines

Grygoriy Zholtkevych

Department of Theoretical and Applied Computer Science,
V.N. Karazin Kharkiv National University
4 Svobody Sqr, 61022, Kharkiv, Ukraine
g.zholtkevych@karazin.ua

Abstract. Modern engineering solutions attract attention of researchers to well-known problems in the field of system theory and cybernetics in general. The realisation problem of a "black box" is one among these problems. In this paper the non-anticipation property for a "black box" is generalised to the case of "black boxes", whose behaviour admits deferred decisions. Furthermore, for such "black boxes" it is shown that they can be realised as pre-machines, which have been introduced by author jointly with his co-authors in series of earlier papers.

Keywords: "black box", deferred responses, sequential processing, pre-machine, transfer function
Key Terms: Computation, Software Component, Specification Process, Mathematical Model

1 Introduction

Let us suppose that two finite alphabets X and Y are given. The first of them we identify as the alphabet of stimuli and the second one as the alphabet of responses. Following to the general cybernetic concept [1, Chapter 6] we can consider any mapping $M\colon X^\omega \to Y^\infty$ as the transfer function of some "black box", whose inputs belong to the set X^ω of infinite sequences of stimuli and outputs belong to the set Y^∞ of finite or infinite sequences of responses. The realisation problem of such a mapping using a machine is the principal problem that is solved by a system engineer. In other words a system engineer transforms a "black" box into a "white box" or "glass box".

The realisation problem had been studied in detail (see, for example, [6]) for the mapping $M\colon X^\omega \to Y^\omega$ holding the following non-anticipation property[1]

$$\begin{aligned}&\text{if } M(\boldsymbol{u}\boldsymbol{x}') = \boldsymbol{y}'\,,\ M(\boldsymbol{u}\boldsymbol{x}'') = \boldsymbol{y}'' \text{ for some finite sequence of stimuli}\\ &\boldsymbol{u} = u_1u_2\dots u_n \text{ and } \boldsymbol{x}', \boldsymbol{x}'' \in X^\omega \text{ then } \boldsymbol{y}' = \boldsymbol{v}\boldsymbol{z}' \text{ and } \boldsymbol{y}'' = \boldsymbol{v}\boldsymbol{z}'' \text{ for}\\ &\text{some finite sequence of responses } \boldsymbol{v} = v_1v_2\dots v_n \text{ and } \boldsymbol{z}', \boldsymbol{z}'' \in Y^\omega\,.\end{aligned} \tag{1}$$

In this case there exists a Moore machine whose transfer function coincides with the mapping M [4, 6].

[1] This property informally means that a "black box" cannot use an information from the future.

We should note the following: the previous formulation for the non-anticipation property implicitly implies that the "black box" responds immediately on each stimulus. However there are systems having another reaction type. It is quite possible such a system behaviour that requires to defer a response for as long as the sufficient amount of the information will be received. For example, complex event processing systems (see [7]) have such a reaction type. Therefore processes of the specification and analysis for such systems require another models or at least models, which generalise already existing ones. This paper is an attempt to solve the realisation problem for "black boxes" with transfer function that satisfies the generalisation being defined below of the non-anticipation property.

2 Prerequisites and Notation

The aim of this section is to give brief survey of some matters and explain the basic notation used below.

At the paper we use the denotation $\mathbb{N}$ for the natural series with 0.

For a set X (it is usually finite) we use the notation:

- X^* denotes the set of all finite sequences (words) whose elements belong to X;
- ε denotes the empty word;
- X^+ denotes the set $X^* \setminus \{\varepsilon\}$;
- X^ω denotes the set of all (infinite) sequences whose elements belong to X;
- X^∞ denotes the union of the sets X^* and X^ω.

Further, we use the denotation $|\boldsymbol{u}|$ for the length of the word $\boldsymbol{u} \in X^*$ and assume that $|\boldsymbol{x}| = +\infty$ for any infinite sequence $\boldsymbol{x} \in X^\omega$.

To refer to the k-th member of a word $\boldsymbol{u} \in X^*$ (or a sequence $x \in X^\omega$) the denotation $u[k]$ (or $x[k]$ respectively) is used.

For a word $\boldsymbol{u} \in X^*$ whose length is equal or greater than n (or a sequence $\boldsymbol{x} \in X^\omega$) by $\boldsymbol{u}[1:n]$ (or $\boldsymbol{x}[1:n]$ respectively) we denote the word $u[1]\dots u[n]$ (or $x[1]\dots x[n]$).

Similarly, for a word $\boldsymbol{u} \in X^*$ whose length is greater than or equal to n (or a sequence $\boldsymbol{x} \in X^\omega$) by $\boldsymbol{u}[n:]$ (or $\boldsymbol{x}[n:]$ respectively) we denote the word $u[n]\dots u[|u|]$ (or the sequence $x[n]x[n+1]\dots$).

3 Non-anticipation Property

In Sec. 1 we have given the definition of the non-anticipation property for a transfer function from X^ω into Y^ω under condition that the corresponding "black box" reacts on each stimulus. Our nearest goal is to generalise the previous definition for the case when a "black box" is capable to decide whether the accumulated information is sufficient for the correct response and generates the response if the decision positive otherwise postpones the response generation.

Firstly, it is needed to say that in this case the class of studied transfer functions are being extended up to the class of mappings from X^ω into Y^∞.

Further, we should specify that the identity of prefixes for streams of stimuli guarantees the identity of prefixes for the corresponding streams of responses. The sequential character of processing streams of stimuli by a "black box" requires that there exists a correspondence between word of stimuli $\boldsymbol{u}$ (as prefix of the corresponding streams) and length $N(\boldsymbol{u})$ of the response word (see Fig. 1).

Fig. 1. A sequential "black box"

The following definition is our attempt to present these considerations as a formal specification.

Definition 1. *We shall say that the non-anticipation property holds for a mapping* $M\colon X^\omega \to Y^\infty$ *if the following is true:*

$$\left.\begin{array}{l} \text{there exists a funtion } N\colon X^* \to \mathbb{N} \text{ such that} \\ \quad 1.\ N(\boldsymbol{u}) \le |\boldsymbol{u}| \text{ for any } \boldsymbol{u} \in X^*; \\ \quad 2.\ \text{if } \boldsymbol{u}' \in X^* \text{ and } \boldsymbol{u}'' = \boldsymbol{u}'x \text{ for some } x \in X \text{ then} \\ \qquad\qquad N(\boldsymbol{u}') \le N(\boldsymbol{u}'') \le N(\boldsymbol{u}') + 1; \\ \quad 3.\ \text{if } \boldsymbol{x}', \boldsymbol{x}'' \in \boldsymbol{u} \cdot X^\omega \text{ for some } \boldsymbol{u} \in X^* \text{ then} \\ \qquad\qquad (M\boldsymbol{x}')[1:\ N(\boldsymbol{u})] = (M\boldsymbol{x}'')[1:\ N(\boldsymbol{u})]; \\ \quad 4.\ \text{for any } \boldsymbol{u} \in X^* \text{ there exist } \boldsymbol{x}', \boldsymbol{x}'' \in \boldsymbol{u} \cdot X^\omega \text{ such that} \\ \qquad\qquad (M\boldsymbol{x}')[1:\ N(\boldsymbol{u})+1] \ne (M\boldsymbol{x}'')[1:\ N(\boldsymbol{u})+1]. \end{array}\right\} \quad (2)$$

Remark 1. Informally, jump points for the function N introduced in Def. 1 determine response instants of the "black box". Items 3) and 4) ensure this interpretation.

Remark 2. Item 2) ensures that the "black box" corresponding to a mapping that holds the non-anticipation property generates at most one response at a stimulus.

Remark 3. Item 3) and 4) of Def. 1 guarantee also that the existence of function N for the mapping $M\colon X^\omega \to Y^\infty$ implies the uniqueness of N.

Remark 4. One can easy see that if $N(\boldsymbol{u}) = |\boldsymbol{u}|$ then Def. 1 and the non-anticipation property given in Sec. 1 specify the same class of mappings.

Now let us consider the partial mapping $\mu\colon X^+ \dashrightarrow Y$ that is defined as follows

$$\mu(\boldsymbol{u})\uparrow \quad \text{iff} \quad N(\boldsymbol{u}) = N(\boldsymbol{u}[1:|\boldsymbol{u}|-1])$$
$$\mu(\boldsymbol{u})\downarrow = M(\boldsymbol{u}\boldsymbol{z})[N(\boldsymbol{u})] \quad \text{iff} \quad N(\boldsymbol{u}) > N(\boldsymbol{u}[1:|\boldsymbol{u}|-1])\,.$$

Item 3) of Def. 1 ensures the uniqueness of determining $\mu(\boldsymbol{u})$.

Remark 5. Returning to Fig. 1, we note that $\mu(\boldsymbol{u}) = y[N(\boldsymbol{u})]$.

To determine the significance of the mapping μ let us consider the following algorithm and proposition.

Require: a sequence of stimuli $\boldsymbol{x} \in X^\omega$
Ensure: to print the corresponding sequence of responses

$n = 1$
while True **:**
  while $\mu(\boldsymbol{x}[1:n])\uparrow$ **:** $n \mathrel{+}= 1$
  $print(\mu(\boldsymbol{x}[1:n]))$
  $n \mathrel{+}= 1$

Algorithm 1: "Black box" algorithm for a mapping $M\colon X^\omega \to Y^\infty$ that holds the non-anticipation property

Proposition 1. *For any $\boldsymbol{x} \in X^\omega$ Algorithm 1 prints the sequence $M\boldsymbol{x}$.*

Proof. Taking into account (2) one can easy see that new response is printed only if $N(\boldsymbol{x}[1:n-1]) \neq N(\boldsymbol{x}[1:n])$. In this case the printed symbol is $(M\boldsymbol{x})[n]$. □

Definition 2. *Let $M\colon X^\omega \to Y^\infty$ be a mapping that holds the non-anticipation property then the corresponding partial mapping $\mu\colon X^+ \dashrightarrow Y$ we shall call its reaction function.*

Conversely, we can consider a partial mapping $\mu\colon X^+ \dashrightarrow Y$ and use Algorithm 1 to define the mapping $M\colon X^\omega \to Y^\infty$.

Proposition 2. *Let $\mu\colon X^+ \dashrightarrow Y$ be a partial mapping then the correspondence $\boldsymbol{x} \in X^\omega \mapsto \boldsymbol{y} \in Y^\infty$, when $\boldsymbol{y}$ is the sequence printed by Algorithm 1 under handling $\boldsymbol{x}$, determines the mapping $M\colon X^\omega \to Y^\infty$ that holds the non-anticipation property.*

Proof. The key idea of the proof consists in the following recursive construction of the function $N\colon X^* \to \mathbb{N}$:

base of recursion: $N(\varepsilon) = 0$;
step of recursion:

$$N(\boldsymbol{u}x) = \begin{cases} N(\boldsymbol{u}), & \text{if } \ \mu(\boldsymbol{x})\uparrow \\ N(\boldsymbol{u}) + 1, & \text{if } \ \mu(\boldsymbol{x})\downarrow \end{cases}.$$

Now, easy seen that the mapping M holds the non-anticipation property. □

4 Automata and Pre-automata

In this section we remind the definition of automata as the simplest discrete systems that respond on external stimuli by changing their states. Automata are actions of free finitely generated monoids on the state sets from the mathematical standpoint. In [3] authors have introduced the notion of a pre-automaton using a generalisation of the notion of an action known as a partial action. Taking into account that these notions is used below and they are not widely used we include this section to give the information necessary for understanding of the further text.

4.1 Automata

We start our consideration reminding the definition of an automaton.

Definition 3. *A triple $\mathcal{A}(X, S_{\mathcal{A}}, \delta_{\mathcal{A}})$ is called an automaton if X is a finite alphabet of stimuli, $S_{\mathcal{A}}$ is a set of states of the automaton, $\delta_{\mathcal{A}}\colon S_{\mathcal{A}} \times X \to S_{\mathcal{A}}$ is a mapping, which is called the transition function of the automaton.*

An automaton behaviour is determined by a right action of the monoid X^* on the state set $S_{\mathcal{A}}$ [2].

Proposition 3. *Let $\mathcal{A}(X, S_{\mathcal{A}}, \delta_{\mathcal{A}})$ be an automaton then the defined recursively defined mapping $\delta^*_{\mathcal{A}}\colon S_{\mathcal{A}} \times X^* \to S_{\mathcal{A}}$*

$$\delta^*_{\mathcal{A}}(s, \varepsilon) = s \quad \text{for any} \quad s \in S_{\mathcal{A}}\,; \tag{3}$$

$$\delta^*_{\mathcal{A}}(s, \boldsymbol{u}x) = \delta_{\mathcal{A}}(\delta^*_{\mathcal{A}}(s, \boldsymbol{u}), x) \quad \text{for} \quad s \in S_{\mathcal{A}},\ \boldsymbol{u} \in X^*,\ x \in X \tag{4}$$

is a right action of monoid X^* on the set $S_{\mathcal{A}}$.

Proof. To prove the proposition it is sufficient to check the equality

$$\delta^*_{\mathcal{A}}(s, \boldsymbol{u}'\boldsymbol{u}'') = \delta^*_{\mathcal{A}}(\delta^*_{\mathcal{A}}(s, \boldsymbol{u}'), \boldsymbol{u}'')$$

for any $s \in S_{\mathcal{A}}$, $\boldsymbol{u}', \boldsymbol{u}'' \in X^*$. Checking is a simple exercise in the application of mathematical induction on the length of $\boldsymbol{u}''$. □

4.2 Pre-automata

The notion of a pre-automaton had been introduced in [3] by replacing the action with the partial action in the definition.

Definition 4. *A triple $\mathcal{P}(X, S_{\mathcal{P}}, \delta^*_{\mathcal{P}})$ is called a pre-automaton if X is a finite alphabet of stimuli, $S_{\mathcal{P}}$ is a set of states of the pre-automaton, $\delta^*_{\mathcal{P}}$ is a right*

partial action of the monoid X^* on the set $S_{\mathcal{P}}$, i.e. it is a partial mapping $\delta^*_{\mathcal{P}}\colon S_{\mathcal{P}} \times X \dashrightarrow S_{\mathcal{P}}$ such that

1. $\delta^*_{\mathcal{P}}(s,\varepsilon)\downarrow = s$ *for all* $s \in S_{\mathcal{P}}$;
2. *if* $\delta^*_{\mathcal{P}}(s,\boldsymbol{u}')\downarrow$ *and* $\delta^*_{\mathcal{P}}(\delta^*_{\mathcal{P}}(s,\boldsymbol{u}'),\boldsymbol{u}'')\downarrow$ *then*

$$\delta^*_{\mathcal{P}}(s,\boldsymbol{u}'\boldsymbol{u}'')\downarrow = \delta^*_{\mathcal{P}}(\delta^*_{\mathcal{P}}(s,\boldsymbol{u}'),\boldsymbol{u}'')\,; \tag{5}$$

3. *if* $\delta^*_{\mathcal{P}}(s,\boldsymbol{u}')\downarrow$ *and* $\delta^*_{\mathcal{P}}(s,\boldsymbol{u}'\boldsymbol{u}'')\downarrow$ *then*

$$\delta^*_{\mathcal{P}}(\delta^*_{\mathcal{P}}(s,\boldsymbol{u}'),\boldsymbol{u}'')\downarrow = \delta^*_{\mathcal{P}}(s,\boldsymbol{u}'\boldsymbol{u}'')\,.$$

4.3 Interrelations between Automata and Pre-automata

In this section we describe some method that allows us to construct a pre-automaton using an automaton.

Suppose that we have taken some automaton $\mathcal{A}(X, S_{\mathcal{A}}, \delta_{\mathcal{A}})$.

Let us define the pre-automaton $\mathcal{P}(X, S_{\mathcal{P}}, \delta^*_{\mathcal{P}})$ in the following manner:

$$\left.\begin{array}{l} 1.\ \text{choose as } S_{\mathcal{P}} \text{ an arbitrary subset of } S_{\mathcal{A}}\,; \\ 2.\ \text{define the partial mapping } \delta^*_{\mathcal{P}}\colon S_{\mathcal{P}} \times X^* \dashrightarrow S_{\mathcal{P}} \text{ as follows for} \\ \quad s \in S_{\mathcal{P}} \text{ and } \boldsymbol{u} \in X^* \text{ let assign that} \\ \qquad \delta^*_{\mathcal{P}}(s,\boldsymbol{u})\uparrow \ \text{ if } \ \delta^*_{\mathcal{A}}(s,\boldsymbol{u}) \notin S_{\mathcal{P}} \ \text{ and} \\ \qquad \delta^*_{\mathcal{P}}(s,\boldsymbol{u})\downarrow = \delta^*_{\mathcal{A}}(s,\boldsymbol{u}) \ \text{ if } \ \delta^*_{\mathcal{A}}(s,\boldsymbol{u}) \in S_{\mathcal{P}}\,. \end{array}\right\} \tag{6}$$

Proposition 4. *The triple defined by construction* (6) *is a pre-automaton.*

Proof. Indeed, item (3) ensures that item 1) of (5) is satisfied.
Further, Prop. 3 implies that items 2) and 3) of (5) are satisfied. □

The assertion just proved demonstrates that the method to obtain pre-automata consists in hiding part of the states.

The converse assertion proved in [3] as Globalisation Theorem ensures that the method considered above is the most general method to obtain pre-automata.

5 Moore Machines and Pre-machines

In this section we discuss the question about how a Moore machine or its generalisation, which we call a Moore pre-machine, can realise a "black box".

5.1 Moore Machines

Usually, automata associate with "black boxes" in the following manner.

Firstly, the class of Moore machines is defined.

Definition 5. *Let $\mathcal{A}(X, S_{\mathcal{A}}, \delta_{\mathcal{A}})$ be an automaton then the corresponding Moore machine is a pentacle $\mathcal{M}_{\mathcal{A}}(X, S_{\mathcal{A}}, \delta_{\mathcal{A}}, s^0_{\mathcal{M}}, \lambda_{\mathcal{M}})$ where $s^0_{\mathcal{M}}$ is some fixed state of $\mathcal{A}$ called the initial state of the machine and $\lambda_{\mathcal{M}}\colon S_{\mathcal{A}} \to Y$ is a mapping called the output function of the machine.*

Then for a Moore machine is determined its reaction function.

Definition 6. *Let $\mathcal{M}_{\mathcal{A}}(X, S_{\mathcal{A}}, \delta_{\mathcal{A}}, s^0_{\mathcal{M}}, \lambda_{\mathcal{M}})$ be a Moore machine then its reaction function $\mu_{\mathcal{M}} \colon X^* \to Y$ is determined by the formula*

$$\mu_{\mathcal{M}}(\boldsymbol{u}) = \lambda_{\mathcal{M}}(\delta_{\mathcal{A}}(s^0_{\mathcal{M}}, \boldsymbol{u}))\,. \tag{7}$$

Finally, we define the transfer function $M_{\mathcal{M}} \colon X^\omega \to Y^\omega$ for the machine $\mathcal{M}_{\mathcal{A}}$ using its reaction function $\mu_{\mathcal{M}}$ and Algorithm 1.

5.2 Moore Pre-machine

Here we repeat all constructions from the previous subsection substituting a pre-automaton for an automaton.

Firstly, the class of Moore pre-machines is defined.

Definition 7. *Let $\mathcal{P}(X, S_{\mathcal{P}}, \delta^*_{\mathcal{P}})$ be a pre-automaton then the corresponding Moore pre-machine is a pentacle $\mathcal{M}_{\mathcal{P}}(X, S_{\mathcal{P}}, \delta^*_{\mathcal{P}}, s^0_{\mathcal{M}}, \lambda_{\mathcal{M}})$ where $s^0_{\mathcal{M}}$ is some fixed state of $\mathcal{P}$ called the initial state of the pre-machine and $\lambda_{\mathcal{M}} \colon S_{\mathcal{P}} \to Y$ is a mapping called the output function of the pre-machine.*

Then for a Moore pre-machine is determined its reaction function.

Definition 8. *Let $\mathcal{M}_{\mathcal{P}}(X, S_{\mathcal{P}}, \delta^*_{\mathcal{P}}, s^0_{\mathcal{M}}, \lambda_{\mathcal{M}})$ be a Moore pre-machine then its reaction function $\mu_{\mathcal{M}} \colon X^+ \dashrightarrow Y$ is determined in the following manner*

$$\begin{array}{ll} 1. & \mu_{\mathcal{M}}(\boldsymbol{u})\uparrow \ \text{ if } \ \delta^*_{\mathcal{M}}(s^0_{\mathcal{M}}, \boldsymbol{u})\uparrow \text{ and} \\ 2. & \mu_{\mathcal{M}}(\boldsymbol{u})\downarrow = \lambda_{\mathcal{M}}(\delta^*_{\mathcal{P}}(s^0_{\mathcal{M}}, \boldsymbol{u})) \ \text{ if } \ \delta^*_{\mathcal{M}}(s^0_{\mathcal{M}}, \boldsymbol{u})\downarrow . \end{array} \tag{8}$$

Finally, we define the transfer function $M_{\mathcal{M}} \colon X^\omega \to Y^\infty$ for the pre-machine $\mathcal{M}_{\mathcal{P}}$ using its reaction function $\mu_{\mathcal{M}}$ and Algorithm 1.

5.3 Posing of Synthesis Problem

The preceding arguments show that machines and pre-machines can be considered as "glass boxes". It is known that machines are "glass boxes" for a proper subclass of the class of all "black boxes" [4, 6]. Therefore we pose the following problem.

Problem 1 (Synthesis Problem). Suppose we have a mapping $M \colon X^\omega \to Y^\infty$ that holds the non-anticipation property.
It is required to describe the properties of the mapping that ensure the existence of a pre-machine $\mathcal{M}_{\mathcal{P}}$ such that $M_{\mathcal{M}} \cong M$.

6 Solving Synthesis Problem

Solving the problems posed at the end of the previous section is given in three stages: firstly, some solution of the problem is constructed, secondly, this solution is reduced, and, finally, the minimality of this reduced solution is proved.

Taking into account the fact that the hypothesis of the problem includes the non-anticipation property for the mapping M we can consider the reaction function μ of the "black box" instead its transfer function.

6.1 Existence of Solution

Thus we assume that two alphabets (the input alphabet X and the output alphabet Y) and a partial mapping $\mu\colon X^+ \dashrightarrow Y$ are given.

Let us choose

$$\left.\begin{array}{l} S_{\mathcal{F}} = X^*\,; \\ \delta_{\mathcal{F}}(\boldsymbol{u}, x) = \boldsymbol{u}x \text{ for } x \in X \text{ and } \boldsymbol{u} \in X^*\,. \end{array}\right\} \tag{9}$$

Now consider the triple $\mathcal{F}(X, S_{\mathcal{F}}, \delta_{\mathcal{F}})$.

Lemma 1. *The triple $\mathcal{F}(X, S_{\mathcal{F}}, \delta_{\mathcal{F}})$ is an automaton such that the right action $\delta^*_{\mathcal{F}}\colon S_{\mathcal{F}} \times X^* \to S_{\mathcal{F}}$ associated with it satisfy the equation*

$$\delta^*_{\mathcal{F}}(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{u}\boldsymbol{v} \tag{10}$$

for all $\boldsymbol{u}, \boldsymbol{v} \in X^*$.

Proof. Checking is reduced to a simple application of the mathematical induction. □

Now let us choose $S^{\mu}_{\mathcal{F}} \subset S_{\mathcal{F}}$ in the following manner:

$$S^{\mu}_{\mathcal{F}} = \{\boldsymbol{u} \in X^* \mid \mu(seq u)\downarrow\} \bigcup \{\varepsilon\}\,.$$

Now applying construction (6) and we obtain the pre-automaton $\mathcal{F}^{\mu}(X, S_{\mathcal{F}}, \delta^{\mu}_{\mathcal{F}})$.

Theorem 1 (Existence of Solutions for the Synthesis Problem). *Let us consider the Moore pre-machine $\mathcal{M}^{\mu}_{\mathcal{F}}(X, S^{\mu}_{\mathcal{F}}, \delta^{\mu\,*}_{\mathcal{F}} s^0_{\mathcal{F}}, \lambda^{\mu}_{\mathcal{F}})$, where*

$$\begin{array}{l} s^0_{\mathcal{F}} = \varepsilon\,; \\ \lambda^{\mu}_{\mathcal{F}}(\boldsymbol{u}) = \mu(\boldsymbol{u}) \quad \textit{if} \quad \mu(\boldsymbol{u})\downarrow\,; \\ \lambda^{\mu}_{\mathcal{F}}(\varepsilon) \quad \textit{is defined arbitrary}\,, \end{array}$$

then $\mu_{\mathcal{M}^{\mu}_{\mathcal{F}}} \cong \mu$.

Proof. Really, $\mathcal{M}^{\mu}_{\mathcal{F}}$ is a Moore pre-machine.
Hence we need to prove that $\mu_{\mathcal{M}^{\mu}_{\mathcal{F}}}(\boldsymbol{u})\downarrow$ if and only if $\mu(\boldsymbol{u})\downarrow$ and the equality $\mu_{\mathcal{M}^{\mu}_{\mathcal{F}}}(\boldsymbol{u}) = \mu(\boldsymbol{u})$ holds on the common domain. But this follows immediately from Lemma 1 and the specification of the pre-machine $\mathcal{M}^{\mu}_{\mathcal{F}}$. □

6.2 Indistinguishability and Syntactic Pre-Machine

The solution that is given in the previous subsection for the Synthesis Problem is too redundant because the state set of the corresponding pre-machine contains too many indistinguishable states. In this subsection we demonstrate the method to eliminate the lack of the construction.

Our consideration refers to some concepts of the theory of ordered sets. The necessary information can be found in [5, Chapter 1].

Definition 9. *We shall say that* $\boldsymbol{u}' \in X^*$ *can not be distinct from* $\boldsymbol{u}'' \in X^*$ *using* μ *(this assertion is below written as* $\boldsymbol{u}' \lesssim_\mu \boldsymbol{u}''$*) if* $\mu(\boldsymbol{u}'\boldsymbol{w})\downarrow$ *implies* $\mu(\boldsymbol{u}''\boldsymbol{w})\downarrow = \mu(\boldsymbol{u}'\boldsymbol{w})$ *for any* $\boldsymbol{w} \in X^*$.

Proposition 5. *The relation "*$\lesssim_\mu$*" is a quasi-order on* X^* *satisfying the following condition: if* $\boldsymbol{u}' \lesssim_\mu \boldsymbol{u}''$ *and* $\boldsymbol{w} \in X^*$ *then* $\boldsymbol{u}'\boldsymbol{w} \lesssim_\mu \boldsymbol{u}''\boldsymbol{w}$.

Proof. Reflexivity and transitivity of the relation is evident.
Now suppose that $\boldsymbol{u}' \in X^*$, $\boldsymbol{u}'' \in X^*$, $\boldsymbol{w} \in X^*$, and $\boldsymbol{u}' \lesssim_\mu \boldsymbol{u}''$.
If $\mu((\boldsymbol{u}'\boldsymbol{w})\boldsymbol{v})\downarrow$ for some $\boldsymbol{v} \in X^*$ then $\boldsymbol{u}' \lesssim_\mu \boldsymbol{u}''$ ensures $\mu(\boldsymbol{u}''(\boldsymbol{w}\boldsymbol{v}))\downarrow = \mu(\boldsymbol{u}'(\boldsymbol{w}\boldsymbol{v}))$ and, therefore, $\mu((\boldsymbol{u}''\boldsymbol{w})\boldsymbol{v})\downarrow = \mu((\boldsymbol{u}'\boldsymbol{w})\boldsymbol{v})$. □

The following simple property of "$\lesssim_\mu$" is used below.

Proposition 6. *The assertions* $\boldsymbol{u}' \lesssim_\mu \boldsymbol{u}''$ *and* $\mu(\boldsymbol{u}')\downarrow$ *imply* $\mu(\boldsymbol{u}'')\downarrow = \mu(\boldsymbol{u}')$.

Proof. To verify the validity of the proposition it is sufficient to put $\boldsymbol{w} = \varepsilon$ in Def. 9. □

Definition 10. *Let* $\boldsymbol{u}', \boldsymbol{u}'' \in X^*$ *then we say that* $\boldsymbol{u}'$ *and* $\boldsymbol{u}''$ *are* μ*-congruent (this assertion is below written as* $\boldsymbol{u} \equiv_\mu \boldsymbol{u}''$*) if both* $\boldsymbol{u}' \lesssim_\mu \boldsymbol{u}''$ *and* $\boldsymbol{u}'' \lesssim_\mu \boldsymbol{u}'$ *are true.*

Proposition 7. *The relation "*$\equiv_\mu$*" is a right congruence on* X^* *that satisfies the following property: if* $\boldsymbol{u}', \boldsymbol{u}'' \in X^*$ *and* $\boldsymbol{u}' \equiv_\mu \boldsymbol{u}''$ *then* $\mu(\boldsymbol{u}')\downarrow$ *if and only if* $\mu(\boldsymbol{u}'')\downarrow$ *and in this case* $\mu(\boldsymbol{u}') = \mu(\boldsymbol{u}'')$.

Proof. The fact that "$\equiv_\mu$" is an equivalence relation follows from the properties of a quasi-order [5, Sec. 1.3]. Its stability relative to the right multiplication follows immediately from the similar property for the relation "$\lesssim_\mu$". Finally, the last assertion of the proposition follows from Prop. 6. □

Let us define

$$\left.\begin{array}{l} S^\mu_{\mathcal{A}} = X^*/\equiv_\mu; \\ \delta^\mu_{\mathcal{A}}([\boldsymbol{u}]_\mu, x) = [\boldsymbol{u}x]_\mu, \end{array}\right\} \quad (11)$$

where $[\cdot]_\mu$ denotes a class of the μ-congruence, $\boldsymbol{u} \in X^*$, and $x \in X$. Note that the property to be a right congruence for the equivalence "$\equiv_\mu$" ensures the correctness of the definition of $\delta^\mu_{\mathcal{A}}$.
Now consider the triple $\mathcal{A}^\mu(X, S^\mu_{\mathcal{A}}, \delta^\mu_{\mathcal{A}})$.

Lemma 2. *The triple* $\mathcal{A}^\mu$ *is an automaton such that the right action associated with it* $\delta^{\mu *}_{\mathcal{A}}\colon S^\mu_{\mathcal{A}} \times X^* \to S^\mu_{\mathcal{A}}$ *satisfies the equation*

$$\delta^{\mu *}_{\mathcal{A}}([\boldsymbol{u}]_\mu, \boldsymbol{v}) = [\boldsymbol{u}\boldsymbol{v}]_\mu \quad (12)$$

for all $\boldsymbol{u}, \boldsymbol{v} \in X^*$.

Proof. The lemma is easy proved by induction on the length of $\boldsymbol{v}$. □

Now we can note that Prop. 7 ensures one of the alternatives:

$$\text{either } \ [\boldsymbol{u}]_\mu \subset \{\boldsymbol{v} \in X^* \mid \mu(\boldsymbol{v})\downarrow\} \ \text{ or } \ [\boldsymbol{u}]_\mu \bigcap \{\boldsymbol{v} \in X^* \mid \mu(\boldsymbol{v})\downarrow\} = \varnothing\,.$$

This remark allows us to choose $S^\mu_{\mathcal{S}} \subset S^\mu_{\mathcal{A}}$ in the following manner:

$$S^\mu_{\mathcal{S}} = \{[\boldsymbol{u}]_\mu \mid \boldsymbol{u} \in X^* \text{ and } \mu(\boldsymbol{u})\downarrow\} \bigcup \{[\varepsilon]_\mu\}\,.$$

Hence we can again apply construction (6) and obtain the pre-automaton $\mathcal{S}^\mu(X, S^\mu_{\mathcal{S}}, \delta^{\mu\,*}_{\mathcal{S}})$.

Theorem 2 (about Syntactic Pre-machine). *Let us consider the Moore pre-machine* $\mathcal{M}^\mu_{\mathcal{S}}(X, S^\mu_{\mathcal{S}}, \delta^{\mu\,*}_{\mathcal{S}}, s^0_{\mathcal{S}}, \lambda^\mu_{\mathcal{S}})$, *where*

$$\begin{aligned} &s^0_{\mathcal{S}} = [\varepsilon]_\mu\,;\\ &\lambda^\mu_{\mathcal{S}}([\boldsymbol{u}]_\mu) = \mu(\boldsymbol{u}) \quad \textit{if} \ \ \mu(\boldsymbol{u})\downarrow;\\ &\lambda^\mu_{\mathcal{S}}([\varepsilon]_\mu) \quad \textit{is defined arbitrary}, \end{aligned}$$

then $\mu_{\mathcal{M}^\mu_{\mathcal{S}}} \cong \mu$.

Proof. Let us note that Prop. 7 ensures the correctness for the definition of the mapping $\lambda^\mu_{\mathcal{S}}$. Further, Lemma 2 guarantees the validity of $\mu_{\mathcal{M}^\mu_{\mathcal{S}}} \cong \mu$. □

Remark 6. We shall call the Moore pre-machine built in the theorem the syntactic pre-machine.

6.3 Syntactic Pre-machine as Minimal Solution of Synthesis Problem

To complete the program indicated above, we need to establish the minimality of the pre-machine $\mathcal{M}^\mu_{\mathcal{S}}$ in any sense.

First of all, we note that the pre-machine $\mathcal{M}^\mu_{\mathcal{S}}$ holds evidently the following property called the reachability: one can obtain any state of the pre-machine applying its partial action to the initial state.

Now let us formulate the main result.

Theorem 3 (about Minimality of Syntactic Pre-machine). *For any reachable Moore pre-machine* $\mathcal{M}_{\mathcal{P}}(X, S_{\mathcal{P}}, \delta^*_{\mathcal{P}}, s^0_{\mathcal{M}}, \lambda_{\mathcal{M}})$ *such that* $\mu_{\mathcal{M}} \cong \mu$ *there exists a mapping* $\psi\colon S_{\mathcal{P}} \to S^\mu_{\mathcal{S}}$ *satisfying the following conditions*

1. *for any* $s \in S_{\mathcal{P}}$ *and* $\boldsymbol{u} \in X^*$ *the assertion* $\delta^*_{\mathcal{P}}(s, \boldsymbol{u})\downarrow$ *implies* $\delta^{\mu\,*}_{\mathcal{S}}(\psi(s), \boldsymbol{u})\downarrow = \psi(\delta_{\mathcal{P}}(s, \boldsymbol{u}))$;
2. $\psi(s^0_{\mathcal{M}}) = s^0_{\mathcal{S}}$;
3. $\mu \cong \mu_{\mathcal{S}^\mu} \circ \psi$.

Proof. The key item of the proof is the construction of the mapping ψ.
Let $s \in S_{\mathcal{P}}$ and $\boldsymbol{u}', \boldsymbol{u}'' \in X^*$ such that $\delta^*_{\mathcal{P}}(s^0_{\mathcal{M}}, \boldsymbol{u}')\downarrow = s$ and $\delta_{\mathcal{P}}(s^0_{\mathcal{M}}, \boldsymbol{u}'')\downarrow = s$ then we can show that $\boldsymbol{u}' \lesssim_\mu \boldsymbol{u}''$.
Indeed, suppose that $\mu(\boldsymbol{u}'\boldsymbol{w})\downarrow$ for some $\boldsymbol{w} \in X^*$. Taking into account that

$\mu_{\mathcal{M}} \cong \mu$ we can write $\mu(\boldsymbol{u}'\boldsymbol{w}) = \lambda_{\mathcal{M}}(\delta^*_{\mathcal{M}}(s^0_{\mathcal{M}}, \boldsymbol{u}'\boldsymbol{w}))$.
Note that the previous equality ensures $\delta^*_{\mathcal{M}}(s^0_{\mathcal{M}}, \boldsymbol{u}'\boldsymbol{w})\downarrow$ and therefore (5) leads to the conclusion that $\delta^*_{\mathcal{P}}(\delta^*_{\mathcal{P}}(s^0_{\mathcal{M}}, \boldsymbol{u}'), \boldsymbol{w})\downarrow$.
Using the supposition $\delta^*_{\mathcal{P}}(s^0_{\mathcal{M}}, \boldsymbol{u}')\downarrow = s$ and (5) we obtain

$$\mu(\boldsymbol{u}'\boldsymbol{w}) = \lambda_{\mathcal{M}}(\delta^*_{\mathcal{P}}(\delta^*_{\mathcal{P}}(s^0_{\mathcal{M}}, \boldsymbol{u}'), \boldsymbol{w})) = \lambda_{\mathcal{M}}(\delta^*_{\mathcal{P}}(s, \boldsymbol{w}))\,.$$

This equation ensures $\delta^*_{\mathcal{P}}(s, \boldsymbol{w})\downarrow$.
Hence the supposition $\delta^*_{\mathcal{P}}(s^0_{\mathcal{M}}, \boldsymbol{u}'')\downarrow = s$ implies $\delta^*_{\mathcal{P}}(\delta^*_{\mathcal{P}}(s^0_{\mathcal{M}}, \boldsymbol{u}''), \boldsymbol{w})\downarrow = \delta^*_{\mathcal{P}}(s, \boldsymbol{w})$.
Thus $\mu(\boldsymbol{u}''\boldsymbol{w})\downarrow = \mu(\boldsymbol{u}'\boldsymbol{w})$ and, therefore, $\boldsymbol{u}'' \lesssim_{\mu} \boldsymbol{u}'$.
Similar reasoning gives $\boldsymbol{u}' \lesssim_{\mu} \boldsymbol{u}''$ and, therefore, $\boldsymbol{u}' \equiv_{\mu} \boldsymbol{u}''$.
Now we can define ψ in the following manner:

$$\psi(s) = [\boldsymbol{u}_s]_{\mu} \quad \text{where} \quad s = \delta^*_{\mathcal{P}}(s^0_{\mathcal{M}}, \boldsymbol{u}_s)\,.$$

Checking the validity of items 1)–3) for the constructed mapping ψ is a simple exercise now. □

7 Conclusion

Thus, we can summarize that the paper gives the algebraic analysis for the problem of realisation "black boxes" by machines. The main results of the analysis are

- the generalisation of the non-anticipation property for "black boxes" that accumulate information for decision;
- the complete solution of the synthesis problem for such "black boxes".

The machines that realise the corresponding transfer functions are based on pre-automata. The class of such algebraic structures had been introduced by author jointly with Prof. M. Dokuchaev and Prof. B. Novikov in earlier papers.

It should be emphasized that issues dealing with computational properties of pre-machines has not considered in the paper. The coverage of these issues requires a separate study.

References

1. Ashby, W.R.: An introduction to cybernetics. Chapman & Hall, London (1956)
2. Clifford, A.H., Preston, G.B.: The Algebraic Theory of Semigroups, Volume 1. AMS (1961)
3. Dokuchaev, M., Novikiov, B., Zholtkevych, G.: Partial actions and automata. Alg. and Discr. Math. 11, 51–63 (2011)
4. Glushkov, V.M.: Some problems in the synthesis of digital automata. USSR Computational Mathematics and Mathematical Physics. 1(3), 399–446 (1962)
5. Harzheim, E.: Ordered Sets. Springer Science+Business Media Inc., New York (2005)

6. Trakhtenbrot, B.A., Barzdin, J.M.: Finite automata: behaviour and synthesis. North-Holland Publishing Company, US (1973)
7. Zholtkevych, G., Novikov, B., Dorozhinsky, V.: Pre-automata and Complex Event Processing. In: Ermolayev, V. et al (eds) ICT in Education, Research, and Industrial Applications. CCIS, vol. 469, pp. 100–116. Springer International Publishing (2014)

An Interleaving Reduction for Reachability Checking in Symbolic Modeling

Alexander Letichevsky[1], Oleksandr Letychevskyi[1], Vladimir Peschanenko[2]

[1]Glushkov Institute of Cybernetics of National Academy of Sciences, Kyiv, Ukraine
{let,lit}@iss.org.ua
[2]Kherson State University, Kherson, Ukraine
vladim@kspu.edu

Abstract. This paper is devoted to the whole problem of interleaving reduction in modeling of concurrent processes. The main notions of insertional modeling were described. The verification problem in terms of insertional modeling was examined. General algorithm of interleaving reduction in terms of insertional modeling was presented. A static and incremental algorithm of reduction for reachability checking was presented. The proof of correctness of presented algorithm was introduced. The results of experiments of such algorithm application was described.

Keywords. Interleaving, predicate transformer, symbolic modeling.

Key Terms. MathematicalModel.

1 Introduction

Usually the multiagent distributed systems are high level non-deterministic. The nature of this non-determinism is symbolic nature of models and concurrency (choice of parallel process which should operate at each time of modeling). One of the main problem of reachability checking in verification is exponential explosion of states number. Some of the sources of such explosion is the number of parallel processes in model and their interleaving[1] .

There are two different approaches for modeling: model checking and symbolic modeling[2]. The model checking tool works with concrete states where state is represented by values of its variables. A transition is occurred by assignment of new values for the variables. The problem of exponential explosion could be solved by using well known model checking methods: methods that introduce partial order to reduce interleaving[3], methods for determining the symmetry when verifying the equivalence of states[4], techniques of abstraction[5], approximation[6], data-flow analyses[7], McMillan's algorithm of unfolding[8].

A state of environment in symbolic modeling presents some formula in corresponded theory (first order logic etc) which covers some set of concrete states. A transition is occurred with a help of predicate transformers (weakest precondition, strongest postcondition[9])[10]. Unfortunately not all methods of model checking for

reducing states space could be applied for symbolic case. The problem which was described previously could be solved with a help of the next symbolic methods: narrowing[11], unfolding concurrent well-structured transition systems[12].

This paper continues the work [13] where an algorithm with some restriction of symbolic model was described. Here we present the algorithm for full symbolic case. The algorithm bases on the McMillan's algorithm adopted to symbolic modeling in notion of insertion modeling [14]. This algorithm bases on notion of permutability which is defined with help of predicate transformer (strongest postcondition, *pt* function below). It was described in [10]. So, the paper is devoted to the solution of the problem of interleaving reduction in insertion models with infinite number of states.

The algebra of behaviors is presented in chapter Behavior Algebras, the verification environments, corresponding insertion function, and predicate transformer are considered in chapter Verification Environments. The normal form of behavior is defined in chapter Behaviors Over Basis *B*. The problem of reachability of the states is described in chapter Verification. The notion of partial unfolding is examined in chapter Partial Unfolding. The optimization of partial unfolding by statically permutable operators is reviewed in chapter Static Permutability Property. The incremental algorithm for reducing of interleaving for transition systems is presented in chapter Main Interleaving Reduction Algorithm. The static algorithm of interleaving reduction is described in chapter Static Interleaving Reduction Algorithm. The statistic of applying of such algorithm to few examples is presented in chapter Examples of Application.

2 Behavior Algebras

One of the main notions of insertion modeling, which is used for describing algorithm of interleaving reduction is behavior algebra. Behavior algebra [14] is a kind of process algebra; it is used to express the behavior of agents (transition systems) considered up to bisimilarity or trace equivalence. To make economic unfolding we need to distinguish sequential and parallel behaviors. So we consider the following modification of the notion of behavior algebra- it is a multisorted algebra with three components: the algebra of *actions*, the algebra of *sequential behaviors*, and the algebra of *parallel behaviors*.

The algebra of sequential behaviors has operations of prefixing: <*action*>.<*sequential behavior*> and one internal operation of nondeterministic choice *(()+())*, which is associative, commutative, and idempotent operation with neutral element 0. We also consider the constant behavior Δ (successful termination), which is a common element of the algebra of sequential and the algebra of parallel behaviors. The operations of action algebra will be considered later.

The algebra of parallel behaviors has the parallel composition *()||()* of sequential behaviors as the main binary operation. It is associative commutative (but is not idempotent) and has the neutral element Δ. It also has the prefixing operation and nondeterministic choice. The algebra of sequential behaviors is implicitly included to the algebra of parallel behaviors by the identity $u = u \| \Delta$ (parallel composition with one component). Unfolding of parallel composition by interleaving will be considered

only after inserting of agents that are formed by parallel composition into the environment.

3 Verification Environments

Verification environments of the form $E = E(U, P, B)$ are defined by the following parameters: the set of *conditional expressions U*, the set of *operators P*, and the set of *basic behaviors B*. The set of conditions and the set of operators are used to define actions (it is a union of these two sets). The set of basic behaviors is used to define the behaviors of agents inserted into environment in the way which will be explained later. We also suppose that some logic language (first order or temporal) called *basic language* is fixed to define the states of environment and checking conditions for verification. The conditional expressions also belong to this language.

The state of environment is represented as $E[u]$, where E is a statement of basic language and u is a parallel composition of sequential behaviors of agents inserted into environment. We suppose that operators are divided into the set of conditional and unconditional operators. Conditional operator has the form $\alpha \to a$ where α is a condition and a is an unconditional operator. Unconditional operator a is identified with conditional operator $1 \to a$. The associative product *()*()* and the function $pt : U \times P \to U$ (predicate transformer) are defined by the set of actions so that the following identities are valid:

$$pt(\alpha, \beta \to a) = pt(\alpha \wedge \beta \to a)$$
$$pt(pt(\alpha, a), b) = pt(\alpha, a * b)$$
$$(\alpha \to a) * (\beta \to b) = pt(pt(\alpha, a) \wedge \beta, b)$$
$$\alpha * \beta = \alpha \wedge \beta$$

Here α and β are conditions, a and b are unconditional operators.

Predicate transformer *pt* is supposed to be monotonic:

$$\alpha \to \beta \Rightarrow pt(\alpha, a) \to pt(\beta, a)$$

In general case, the *pt* function is defined by some concrete syntax. An example of such pair *(syntax, pt)* can be found in [16].

Example. The basic language is a first order language. Conditions are formulae over simple attributes - symbols that change their values when a system changes its state. Formally they are considered as function symbols with arity 0. Unconditional operators are assignments (parallel assignments, sequences of assignments, if-then-else operators, loops with finite number of repetitions, etc.). As usually in this case,

$$pt(\alpha(x), (x_1 := t_1(x), x_2 := t_2(x), \ldots)) = \exists z(\alpha(z) \wedge (x_1 = t_1(z) \wedge x_2 = t_2(z) \wedge \ldots))$$

Actually this is the strongest postcondition for precondition α.

Example of conditional operator. Let x be an integer variable, $u = (x < 5) \to (x := x + 1)$ be an operator, $x > 3$ is statement in basic language, $u \,||\, u$ is a behavior. For this case, $U = \not\subset, P = \{u\}, B = \{u = (x < 5) \to (x := x + 1)\}$. The equation $u = (x < 5) \to (x := x + 1)$ considered here as a basic behavior and it used for definition of agent behavior $u \,||\, u$.

In insertion modeling environment considered as agent with insertion function. So, **Insertion function** is defined by the following identities and rules of operational semantics.

1. $E[u,v] = E[u \,||\, v]$, u,v are agents with sequential behavior (see sec. 1).

Identities for conditions.

2. $E[\alpha.u + v] = E[v]$, if $(E \wedge \alpha) = 0$.

3. $E[\alpha.\beta.u + v] = E[\alpha \wedge \beta.u + v]$, if $(E \wedge \alpha) \neq 0$ (merging conditions).

4. $E[\alpha.\beta \to a.u + v] = E[\alpha \wedge \beta \to a.u + v]$, if $(E \wedge \alpha \wedge \beta) \neq 0$. Special cases of these identities are obtained when v=0 or $\beta = 1$.

5. $E[\alpha.\varepsilon] = E \wedge \alpha[\varepsilon]$, if $(E \wedge \alpha) = 0$.

Identities for operators.

6. $E[a.u + v] = E[v]$, if $pt(E,a) = 0$.

7. $E[a.u] = a.pt(E,a)[u \,||\, \varphi(a,E)]$, if $pt(E,a) \neq 0$, $\varphi(a,E)$ is a parallel composition of sequential behaviors (it generates some new parallel branches). If $\varphi(a,E) = \Delta$, then $u \,||\, \varphi(a,E) = u \,||\, \Delta = u$ and u remains unchanged.

Nondeterministic choice.

8. $E[a.u + a.v + w] = E[a.(u + v) = w]$. The use of left distributivity means that environment considers behavior expressions up to trace equivalence. It also means that a system uses delayed (angelic) choice.

9. $E[u + \Delta] = E[u] + E[\Delta]$. The states $E[0]$ and $E[\Delta]$ are called *terminal states of the environment*. Formally, the states of the form $E[0]$ are equivalent to 0, and states of the form $E[\Delta]$ are equivalent to Δ (if $E[\Delta] = E[\Delta] + \Delta$ is added). But from the point of view of verification it is useful to distinguish syntactically different terminal states.

Parallel behaviors.

10. $E[u] = E[v] \Rightarrow E[u \,||\, w] = E[v \,||\, w]$. Therefore all identities for conditions and operators can be applied within the parallel composition. A component $a_1.u_1 + \ldots + a_n.u_n$ of parallel composition is called *degenerated relative to the state E*, if for all operators $a_i.pt(E,a_i) = 0$ and for all conditions α_i it is true that $(E \wedge \alpha_i) = 0$. Each component that is degenerated relatively to the state E is equivalent to 0 relatively to this state.

11. $E[u] + F[v] = F[v]$, if parallel composition u contains degenerated component relative to E. So all states of environment with degenerated components are equivalent to 0.

12. $E[u + \Delta \,||\, v] = E[u \,||\, v] + E[v]$.

13. $E[a_1.u_1 + a_2.u_2 + \ldots] = E[a_1.u_1 \,||\, v] + E[a_2.u_2 \,||\, v] + \ldots$, if all actions a_i are different, if a_i is a condition then u_i is terminal constant, and v does not contain components degenerated relatively to the state E. The state of environment $E[u]$ is called *dead lock state*, if there are no transitions from this state, but u is not a successful termination. If there is at least one degenerated component in parallel

composition, then the corresponding state is a dead lock state. All dead lock states are equivalent to 0, but it is useful to distinguish them as well as terminal constants. The rules (9), (12), and (13) are called *unfolding of nondeterministic choice*.

14. $E[a_1.u_1 \,||\ldots||\, a_n.u_n] = \sum_{i=1}^{n} a_i.(\ldots||\, a_{i-1}.u_{i-1} \,||\, u_i \,||\, a_{i+1}.u_{i+1} \,||\ldots)$, if all components of parallel composition are non-degenerated. This relation is called a *full unfolding algorithm for a parallel composition*. This is a complete unfolding and the main result of this chapter shows that it is not needed to make the complete unfolding at each step of verification. Let $u = a_1.u_1 \,||\ldots||\, a_n.u_n$,

$$unfold(u,i) = a_i.(\ldots||\, a_{i-1}.u_{i-1} \,||\, u_i \,||\, a_{i+1}.u_{i+1} \,||\ldots)$$

then identity (14) can be rewritten as

14a. $E[a_1.u_1 \,||\ldots||\, a_n.u_n] = \sum_{i=1}^{n} unfold(u,i)$.

Environment does not distinguish trace equivalent behaviors and consequently, bisimilar states of environment are trace equivalent[14]. The identity (14) defines the main transition rule for the system:

$$E[u] \xrightarrow{a_i} E'[u'],$$

if u is a parallel composition with non-degenerated components and $E'[u']$ is defined by the identity (7).

4 Behaviors over Basis B

The set of symbols is given for the set B of behavior basis. These symbols are called *basic sequential behaviors*. The expression of the algebra of sequential behaviors constructed from these symbols and termination constants is called *sequential behavior over basis B*. Suppose that for each symbol $v \in B$ an equation of the form $v = F_v(v_1, v_2, \ldots)$ is given with sequential behavior over basis B as a right hand side. This equation is called the *definition of a basic behavior v*. The application of this definition (the substitution of the left hand side by the right hand one) is called the *unfolding of this behavior*. System of basic behaviors is called non-degenerated if each path in the tree representation of the expression $v = F_v(v_1, v_2, \ldots)$ contains at least one operator.

Normal form of sequential behavior is an expression of the form $a_1.u_1 + a_2.u_2 + \ldots + a_n.u_n + \varepsilon$ where $u_1, u_2, \ldots$ are sequential behaviors. If a_i is a condition, then u_i is a termination constant, $n \geq 0$, and all actions are different (not equivalent with respect to the environment E), because of delayed (angelic) choice (see sec. 2).

Each sequential behavior u over non-degenerated basis in a state $E[u]$ can be reduced to a normal form v equivalent to u with respect to E.

Parallel behavior over B is a parallel composition of sequential behaviors over B.

Normal form of parallel behavior is a nondeterministic sum of behaviors of the form $a_1.u_1 + a_2.u_2 + \ldots$, where $u_1, u_2, \ldots$ are sequential behaviors over B, $a_1, a_2, \ldots$

are operators or conditions such that if a_i is a condition, then u_i is a termination constant.

Normal form of environment state is a term of the form $\sum_{i\in I} a_i.E_i[u_i]+\sum_{j\in J}\Delta$ or 0. *Each environment state with non-degenerated system of basic behaviors is a trace equivalent to some normal form.*

5 Verification

A property ξ of environment state is said to be *correct* if it does not distinguish equivalent states. A property ξ of environment state is monotonic if $E \to E' \Rightarrow \xi(E[u]) \to \xi(E'[u])$.

5.1 Verification problem in terms of insertion modeling

Let S_1, S_2 be state of the model M. The problem of reachability checking is the answer to the question if a path exists from the state S_1 to the state S_2 on model M, or not. Usually models are highly non-deterministic. This non-determinism is based on interleaving of parallel processes: $a||b=(a;b)+(b;a)$ (here a,b are some processes, "||" is parallel composition, ";" is sequential composition and "+" is non-deterministic composition). From other side this non-determinism could produce additional paths from S_1 to S_2 and additional states. So, let call interleaving reduction problem an answer to the question how to reduce non-determinism of the model M to find the path from S_1 to S_2 as quickly as possible.

For a given set Ξ of correct and monotonic checked properties, defined on the set of environment states, the set of initial states defines which properties are reachable (not reachable) from the initial states for a finite number of steps or a number of steps bounded by some constant.

It is supposed that the set of properties to be checked contains the property of a state "to be a dead lock" and a property "to be a state of successful termination".

The simplest verification algorithm is exhaustive unfolding of initial states up to saturation or depletion of a given number of steps. It uses the following formula of unfolding: $\sum_{i=1}^{n} E[unfold(u,i)]$. Such algorithm was described in [14]. It builds all states space for reachability checking which isn't possible always. The properties to be checked are checked in the process of unfolding and the states that satisfy checked properties are collected. More economic unfolding algorithm can be constructed using the following *partial unfolding algorithm.*

6 Partial Unfolding

Two operators a and a' are called permutable regarding the state of E if $E[a*a']=E[a'*a]$ and dynamically permutatable regarding the state E (denoted by

$a \xleftrightarrow{E} a'$) if $E[a*a'] = E[a'*a] \neq 0$. Let $E[u] = E[a_1.u_1 \,||\ldots||\, a_n.u_n]$ is a state of the environment. Let's select the component $s = a_i.u_i$ and build $nonp(E,a_i) = \{a_j \mid i \neq j \wedge \neg(a_i \xrightarrow{E} a_j)\}$. We obtain:

$$punfold(E,u,i) = A(i) + B(E,i) + C(E,i)$$

$$A(i) = a_i.(\ldots||a_{i-1}.u_{i-1}\,||u_i\,||a_{i+1}.u_{i+1}\,||\ldots)$$

$$B(E,i) = \sum_{i\neq j\wedge(a_i,a_j)\in nonp(E,s)} a_j.(\ldots||a_{j-1}.u_{j-1}\,||u_j\,||a_{j+1}.u_{j+1}\,||\ldots)$$

$$C(E,i) = \sum_{k\neq i\wedge(a_k,a_i)\notin nonp(E,a_i)\wedge a_k \xleftrightarrow{E} a_w} a_k.(\ldots||a_{k-1}.u_{k-1}\,||((p;a_w);u'_k)\,||a_{k+1}.u_{k+1}\,||\ldots)$$

In the last formula $((p;a_w);u'_k) = u_k$ and *p* are sequences of compositions of actions (behavior). Function *punfold* is called *partial unfolding of parallel composition*. Let's consider the following algorithm of reachability checking: we need to check the properties on a current state of the environment and each state that is reachable from this in one step. Partial unfolding is used for main function of unfolding states. This algorithm is called partial unfolding algorithm of reachability checking.

In general, the *punfold* uses the notion of dynamic permutability of operators, but it is not optimal, because it uses 4 times application of function predicate transformer *pt* for each pair of operators. Using *punfold* can be optimized by using the concept of static permutability of operators. Algorithm which uses *punfold* with some optimization is considered in section 6.3.

6.1. Optimization of partial unfolding of states.

Theorem 1. If two operators $p = \alpha \to a, q = \beta \to b$ are permutable regarding the states $E_1 = \alpha \wedge \beta$, $E_2 = \neg\alpha \wedge \beta$, $E_3 = \alpha \wedge \neg\beta$ then they are permutable regarding any state [13].

The sufficient condition of permutability of two operators $p = \alpha \to a, q = \beta \to b$ is valid under the following conditions:

1. $pt(\alpha \wedge pt(\alpha \wedge \beta, b), a) = pt(\beta \wedge pt(\alpha \wedge \beta, a), b)$;
2. $pt(\alpha \wedge pt(\neg\alpha \wedge \beta, b), a) = 0$;
3. $pt(\beta \wedge pt(\alpha \wedge \neg\beta, a), b) = 0$.

Example 1. Let *a,b:int* and $1[init.(a_1.good\,||b_0.bad + b_1.good)]$ is initial state and behavior, where $init, a_1$, b_0, b_1 - operators. Agent's behavior could be represented by the following list of equations: $init = ((a = b) \to 1).AndFork$,

$AndFork = a_1\,||(b_0 + b_1)$, $a_1 = ((a = 1) \to 1)$, $b_0 = ((b = 0) \to 1)$, $b_1 = ((b = 1) \to 1)$.

Sufficient condition of permutability for operators a_1, b_0, b_1 is performed in this case, but there can be a case in the simulation where the state of the environment includes some formula, which combines predicate memory of various parallel processes *(a=b)*. So, one of the operator will not be applicable, ie a pair of operators

will be dynamically permutable regarding this state. Thus, the notion of sufficient conditions of permutability of operators need to be strengthened.

To improve the usage of permutability for this example, we need $sat(E \wedge \alpha \wedge \beta) = 1$, otherwise operators will be dynamically permutable regarding state *E*. Let's try to obtain a sufficient condition for dynamic permutability of two operators regarding some condition *E*.

The notion of dynamic permutability of two operators *p, q* regarding some state *E* uses a condition:

$$E[p*q] = E[q*p] \neq 0$$

So, let $E' = pt(\alpha \wedge pt(\alpha \wedge \beta, b), a) = pt(\beta \wedge pt(\alpha \wedge \beta, a), b) \neq 0$ and try to apply backward predicate transformer to the state *E*. We obtain:

$$pt^{-1}(pt^{-1}(E', \beta, b), \alpha, a) = E''_{(q,p)}, pt^{-1}(pt(E', \alpha, a), \beta, b) = E''_{(p,q)}.$$

Theorem 2. If $E' = pt(\alpha \wedge pt(\alpha \wedge \beta, b), a) = pt(\beta \wedge pt(\alpha \wedge \beta, a), b) \neq 0$ then $E''_{(q,p)} \wedge E''_{(p,q)} \neq 0$.

Proof.

Let's assume the contrary that $E''_{(q,p)} \wedge E''_{(p,q)} = 0$. Since the backward predicate transformer turns back to its possible state transition set, it means that $(\alpha \wedge \beta \to E''_{(q,p)}) \wedge (\alpha \wedge \beta \to E''_{(p,q)})$. State $E''_{(p,q)}$ ($E''_{(q,p)}$) specifies a set of concrete states from which transitions from state $\alpha \wedge \beta$ with operators *p* and *q* (*q* and *p*) exist, which means that $\alpha \wedge \beta \wedge E''_{(p,q)} \wedge \alpha \wedge \beta \wedge E''_{(q,p)} \Rightarrow \alpha \wedge \beta \wedge E''_{(p,q)} \wedge E''_{(q,p)} \neq 0$. So, we got a contradiction, because if $E''_{(q,p)} \wedge E''_{(p,q)} = 0$ then $E' = 0$. The theorem is proved.

This condition means that if two operators were dynamically permutable regarding *E* then it is necessary that current state of the environment should satisfy theorem 2.

Let *E* be some state of environment.

Theorem 3. If two operators $p = \alpha \to a, q = \beta \to b$ satisfy the sufficient condition of permutability and $E \wedge E''_{(q,p)} \wedge E''_{(p,q)} \neq 0$ then $E[p*q] = E[q*p] \neq 0$.

Proof.

Let's consider the condition of dynamic permutability regarding *E*: $E[p*q] = E[q*p] \neq 0$.

$$E[p*q] = E[(\alpha \to a)*(\beta \to b)] \Rightarrow pt(E \wedge \alpha, a)[\beta \to b] \Rightarrow$$
$$\Rightarrow pt(\alpha \wedge pt(E \wedge \beta, b), a) \Rightarrow pt(\alpha \wedge pt(E \wedge \beta \wedge (\alpha \vee \neg\alpha), b), a) \Rightarrow$$
$$\Rightarrow pt(\beta \wedge pt(E \wedge \alpha \wedge \beta \vee E \wedge \alpha \wedge \neg\beta, a), b) \Rightarrow$$
$$\Rightarrow pt(\beta \wedge pt(E \wedge \alpha \wedge \beta), a) \vee \beta \wedge pt(E \wedge \alpha \wedge \neg\beta), a), b) \Rightarrow$$
$$\Rightarrow pt(\beta \wedge pt(E \wedge \alpha \wedge \beta), a), b) \vee pt(\beta \wedge pt(E \wedge \alpha \wedge \neg\beta), a), b)$$

Next, let's consider in details the sufficient condition permutability of operators that satisfies the operators *p, q*:

$$E[p*q] = E[(\alpha \to a)*(\beta \to b)] \Rightarrow pt(E \wedge \alpha, a)[\beta \to b] \Rightarrow$$
$$\Rightarrow pt(\beta \wedge pt(E \wedge \alpha, a), b) \Rightarrow pt(\beta \wedge pt(E \wedge \alpha \wedge (\beta \vee \neg\beta), a), b) \Rightarrow$$
$$\Rightarrow pt(\beta \wedge pt(\alpha \wedge \neg\beta \wedge E, a) \vee \beta \wedge pt(\alpha \wedge \neg\beta \wedge \neg E, a), b) = 0 \Rightarrow$$

$$\Rightarrow pt(\alpha \wedge pt(\neg\alpha \wedge \beta \wedge E, b), a) \vee pt(\alpha \wedge pt(\neg\alpha \wedge \beta \wedge \neg E, b), a) = 0 \Rightarrow$$

$$\Rightarrow pt(\beta \wedge pt(\alpha \wedge \neg\beta \wedge E, a), b) = 0 \wedge pt(\beta \wedge pt(\alpha \wedge \neg\beta \wedge \neg E, a), b) = 0$$

Equality $E[p*q] = E[q*p]$ shall be satisfied because otherwise the operators *p*, *q* do not satisfy the sufficient condition permutability of operators (Theorem 1). Thus, we have:

$$pt(\beta \wedge pt(E \wedge \alpha \wedge \beta), a), b) \vee pt(\beta \wedge pt(E \wedge \alpha \wedge \neg\beta), a), b) =$$

$$= pt(\alpha \wedge pt(E \wedge \beta \wedge \alpha, b), a) \vee pt(\alpha \wedge pt(E \wedge \beta \wedge \neg\alpha, b), a) \Rightarrow$$

$$\Rightarrow pt(\beta \wedge pt(E \wedge \alpha \wedge \beta), a), b) = pt(\alpha \wedge pt(E \wedge \beta \wedge \alpha, b), a)$$

Let's consider opposite:

$$pt(\beta \wedge pt(E \wedge \alpha \wedge \beta), a), b) = pt(\alpha \wedge pt(E \wedge \beta \wedge \alpha, b), a) = 0$$

Let's continue to consider sufficient conditions of permutability:

$$pt(\alpha \wedge pt(\alpha \wedge \beta, b), a) = pt(\beta \wedge pt(\alpha \wedge \beta, a), b) \neq 0 \Rightarrow$$

$$\Rightarrow pt(\alpha \wedge pt(\alpha \wedge \beta \wedge (E \vee \neg E), b), a) = pt(\beta \wedge pt(\alpha \wedge \beta \wedge (E \vee \neg E), a), b) \neq 0 \Rightarrow$$

$$\Rightarrow pt(\alpha \wedge pt(\alpha \wedge \beta \wedge E \vee \alpha \wedge \beta \wedge \neg E), b), a) =$$

$$= pt(\beta \wedge pt(\alpha \wedge \beta \wedge E \vee \alpha \wedge \beta \wedge \neg E), a), b) \neq 0 \Rightarrow$$

$$\Rightarrow pt(\alpha \wedge pt(\alpha \wedge \beta \wedge E), b), a) \vee pt(\alpha \wedge pt(\alpha \wedge \beta \wedge \neg E), b), a) =$$

$$= pt(\beta \wedge pt(\alpha \wedge \beta \wedge E), a), b) \vee pt(\beta \wedge pt(\alpha \wedge \beta \wedge \neg E), a), b) \neq 0 \Rightarrow$$

$$\Rightarrow pt(\alpha \wedge pt(\alpha \wedge \beta \wedge \neg E), b), a) = pt(\beta \wedge pt(\alpha \wedge \beta \wedge \neg E), a), b) \neq 0$$

This means that the condition $\alpha \wedge \beta \wedge \neg E \wedge E''_{(q,p)} \wedge E''_{(p,q)} \neq 0$ should be satisfied. But we have the following condition $E \wedge E''_{(q,p)} \wedge E''_{(p,q)} \neq 0$. Thus, both conditions must be satisfied, however:

$$\alpha \wedge \beta \wedge \neg E \wedge E''_{(q,p)} \wedge E''_{(p,q)} \wedge E \wedge E''_{(q,p)} \wedge E''_{(p,q)} = 0$$

So we got a contradiction. The theorem is proved.

If there are two operators $p = \alpha \to a, q = \beta \to b$ that satisfy the sufficient condition of permutability. Condition $E \wedge E''_{(q,p)} \wedge E''_{(p,q)} \neq 0$ is called sufficient condition of dynamic permutability of operators *p, q* regarding the environment *E*.

From a practical point of view, let's try to identify requirements for operators with which we can determine statistically whether they satisfy the sufficient condition of dynamic permutability or not.

Let *E* be a state of the environment, and *p* - an operator. The set *A(E)* is called the *set of all attributes from state E* and *A(p) is* called the *set of all attributes in the statement p*[15].

Two operators $p = \alpha \to a, q = \beta \to b$ are called *statically permutable* if they satisfy the following conditions:

$$A(p) \cap A(q) = \varnothing \wedge pt(\alpha, a) \neq 0 \wedge pt(\beta, b) \neq 0$$

Theorem 4. If two operators $p = \alpha \to a, q = \beta \to b$ are statically permutable then they are dynamically permutable.

Proof.

To prove the theorem we need to show that these operators satisfy necessary condition of permutability of operators in this case.

Since $A(p) \cap A(q) = \varnothing \wedge pt(\alpha,a) \neq 0 \wedge pt(\beta,b) \neq 0$ and *theorem 1* then

$pt(\beta \wedge pt(\alpha \wedge \neg\beta, a), b) = pt(\beta \wedge \neg\beta \wedge pt(\alpha, a), b) = 0$

$pt(\alpha \wedge pt(\neg\alpha \wedge \beta, b), a) = pt(\alpha \wedge \neg\alpha \wedge pt(\beta, b)) = 0$

$pt(\alpha \wedge pt(\alpha \wedge \beta, b), a) = pt(\beta \wedge pt(\alpha \wedge \beta, a), b) \Rightarrow$

$\Rightarrow pt(\alpha \wedge \alpha \wedge pt(\beta, b), a) = pt(\beta \wedge \beta \wedge pt(\alpha, a), b) \Rightarrow$

$\Rightarrow pt(\alpha \wedge pt(\beta, b), a) = pt(\beta \wedge pt(\alpha, a), b) \Rightarrow$

$\Rightarrow pt(\alpha, a) \wedge pt(\beta, b) = pt(\alpha, a) \wedge pt(\beta, b)$

The theorem is proved.

This theorem means that if a predicate that combines memory in a state of environment with different operators is absent then checking the necessary condition of dynamic permutability is not required. Since in this case a usage of one of these operators does not affect the applicability of another operator. The appearance and disappearance of such predicates can be defined statically and syntactically.

Thus, in *Example 1* operators are statically permutable, but after applying *init* operator formula will contain predicate that combines memory of operators a_1, b_0 and a_1, b_1. So, we have to use sufficient condition for dynamic permutability of pairs of operators, a_1, b_0 and a_1, b_1 regarding the state of the environment after application of *init* operator . So, $E = (a = b)$. Let's statically compute sufficient condition of permutability of operators:

$(a_1, b_o) : E''_{(a_1,b_0)} \wedge E''_{(b_0,a_1)} = (a = 1) \wedge (b = 0)$

$(a_1, b_1) : E''_{(a_1,b_1)} \wedge E''_{(b_1,a_1)} = (a = 1) \wedge (b = 1)$

Next let's try to apply sufficient condition of dynamic permutability of operators regarding the condition *E* for both pairs of operators:

$(a_1, b_o) : E \wedge E''_{(a_1,b_0)} \wedge E''_{(b_0,a_1)} = (a = b) \wedge (a = 1) \wedge (b = 0) = 0$

$(a_1, b_1) : E \wedge E''_{(a_1,b_1)} \wedge E''_{(b_1,a_1)} = (a = b) \wedge (a = 1) \wedge (b = 1) \neq 0$

Thus, operators (a_1, b_o) will be dynamically permutable regarding the condition *E*, and operators (a_1, b_1) will be dynamically permutable. This means that interleaving will be removed in correct way for this problem.

6.2. The Problem of Reachability of Some State

The approach proposed in the previous sections can be applied to the problem of finding deadlocks in a given model, but if the user specifies a state of environment you want to check coverage, whereas previously proposed approach should be strengthened.

Example 2. Let *a,b:int* and 1[*init.*($a_1 || b_1$)] be initial behavior and a state of the environment, where *init*, a_1, b_1 - operators. Agent's behavior could be represented by the following list of equations:

$$init = ((a = 0) \wedge (b = 0) \to 1), AndFork, AndFork = a_1 || b_1,$$
$$a_1 = (1 \to (a := 1)), b_1 = (1 \to (b := 1))$$

Let's check recheability of the state $F = (a = 0) \wedge (b = 1)$.

After applying the operator *init* obtains the state of the environment $E = (a = 0) \wedge (b = 0)$. Operators a_1, b_1 are statically permutable and can be applied to the state *E*, which means that they are dynamically permutable regarding *E*. So, $E[a_1 \,||\, b_1] \Rightarrow E[a_1.b_1]$, which means that the operator b_1 never will be applied before the operator a_1 and user defined state *F* will be unreachable after interleaving reduction. Let's try to enhance sufficient condition of operators permutability regarding some state *E* with some conditions related to formula *F*. $F \wedge (E[a_1 * b_1]) = F \wedge (E[b_1 * a_1]) = 0$ for this example then we consider conditions for operators separately (not for pairs of operators).

Let $p = \alpha \to a$ be an operator.

Theorem 5. If $\alpha \wedge pt^{-1}(pt(\alpha, a), \alpha, a) \neq 0$ then $pt(\alpha, a) \neq 0$.

Proof.

Let's consider the opposite $\alpha \wedge pt^{-1}(pt(\alpha, a), \alpha, a) \neq 0$ and $pt(\alpha, a) = 0$. In this case by performed substitution it can be easily obtained the following:

$$\alpha \wedge pt^{-1}(pt(\alpha, a), \alpha, a) \neq 0 \Rightarrow \alpha \wedge pt^{-1}(0, \alpha, a) \neq 0 \Rightarrow 0 \neq 0$$

So we got a contradiction. The theorem is proved.

The operator $p = \alpha \to a$ is called permutable regarding some user defined state *F*, if the following conditions are satisfied:

1) $\alpha \wedge pt^{-1}(pt(\alpha, a), \alpha, a) \neq 0$;

2) $F \wedge \alpha \wedge pt^{-1}(pt(\alpha, a), \alpha, a) = F \wedge pt(\alpha, b)$.

This permutability means that an operator does not change the state of the environment in order to reach the user defined state changed. From reachability point of view we are interested in two cases (if $\alpha \wedge pt^{-1}(pt(\alpha, a), \alpha, a) \neq 0$):

1) $F \wedge \alpha \wedge pt^{-1}(pt(\alpha, a), \alpha, a) \neq 0 \wedge F \wedge pt(\alpha, b) = 0$;

2) $F \wedge \alpha \wedge pt^{-1}(pt(\alpha, a), \alpha, a) = 0 \wedge F \wedge pt(\alpha, b) \neq 0$.

In first case, the reachability of user defined state should be checked immediately before application of an operator, and in the second case - after.

If operators satisfy the sufficient condition of dynamic permutability, but at least one of them is not permutable regarding a user defined state then this operator should be applied first.

This approach can be applied to any algorithm of unfolding.

So, for checking of reachability of the user defined state *F* the notion of permutability regarding the user defined state could be used. You can't consider a pair of operators if both of them do not satisfy this condition.

Coming back to example 2. Operator a_1 will not be permutable regarding the user defined state *F*:

$1 \wedge pt^{-1}(pt(1, a := 1), 1, a := 1) \Rightarrow 1 = E_1$

$pt(1, a := 1) \Rightarrow (a = 1) = E_2$

$F \wedge E_1 = F \wedge E_2 \Rightarrow (a = 0) \wedge (b = 1) \wedge 1 = (a = 0) \wedge (b = 1) \wedge (a = 1) \Rightarrow 0$

Operator b_1 is permutable regarding *F*:

$1 \wedge pt^{-1}(pt(1, b := 1), 1, b := 1) \Rightarrow 1 = E_1$

$pt(1, b := 1) \Rightarrow (b = 1) = E_2$

$F \wedge E_1 = F \wedge E_2 \Rightarrow (a = 0) \wedge (b = 1) \wedge 1 = (a = 0) \wedge (b = 1) \wedge (b = 1) \Rightarrow 1$

From other side operators a_1, b_1 are statically permutable since $E = (a = 0) \wedge (b = 0)$ has no predicates that combine memory of these operators.

This means that in this case you should first apply an operator b_1, then you need to check reachability of F (since $F \wedge E_1 \Rightarrow (a = 0) \wedge (b = 1) \wedge 1 \neq 0$) before applying a_1. And after that you can try to apply a_1. So,

$pt((a = 0) \wedge (b = 0), b := 1) = (a = 0) \wedge (b = 1)$

Then let's check the reachability of user defined state:

$(a = 0) \wedge (b = 1) \wedge (a = 0) \wedge (b = 1) \Rightarrow 1$

So, reachability of user defined state is proved.

6.3. The Main Interleaving Reduction Algorithm

Let $E[u]$ be a model (an initial state of the environment and behavior), where u is behavior, and F is some user defined state which reachability should be checked. So, we need to check reachability of F in the model $E[u]$ and all its deadlocks.

The main interleaving reduction algorithm for reachability checking is represented in fig. 1.

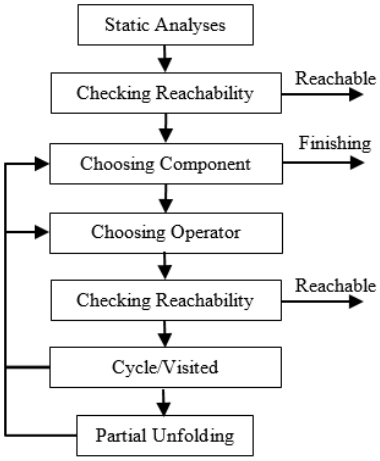


Fig. 1. The main interleaving reduction algorithm for reachability checking

Static Analyses. In the initial behavior u we look for set Op_n (set of operators of a n-th parallel process) on each parallel process. Next, for each pair of operators from different parallel processes we build a table: $H : N \times Op_N \rightarrow N \times Op_N \times G \times Bool$. This table by a pair (parallel process identifier and operator) returns four (number of parallel process that is not equal to the previous one, and the operator which is

permutable to current one, the last two parameters are sufficient condition for permutability of operators and flag for static permutability of operators).

For each pair of operators that does not satisfy the sufficient condition of permutability of operators the table is filled: $D: N \times Op_N \to \{N, Op_N\}$, where $\{N, Op_N\}$ is a set of pairs: the number of parallel process and an operator, which does not satisfy the sufficient condition of permutability.

Each operator is constructed $Flt: Op_N \to Bool \times Bool \times Bool$, which defines the triple for each operator in set: value of reachability of user defined state before and after application of an operator, the third value is 1 if all operators from other processes are statically permutable with this one, and 0 if not.

Checking Reachability. Checking reachability of user defined state *F*. If the filter is reachable then saving corresponded trace and stop modeling.

Choosing Component. We build normal form (*section 4*). From list of components we should choose one to continue working. Here we propose to select first component from the list, but in general case here some heuristics could be applied (it's out of scope of this paper). If there is no component left then finishing.

Choosing Operator. In a chosen component we select operators in the following order. First we check the applicability of operators for which the third option from the table *Fpl* is 1. If there are no operators or they can't be applied, then we choose other operators. If the flag state of the reachability of user defined state is 1 then we first try to apply such operators that are permutable regarding a user defined condition for both operators in the table *Flt* being 1. If the flag state of user defined state equals 0 then we choose to consider operators whose value pairs in the table *Flt* are (0,1). If one of such operators is applicable then after his application we need to check the reachability of user defined state. If user defined state is reachable then finishing. If no such operators left then deadlock is obtained and we get new component. Otherwise, finishing.

Cycle/Visited. Checking cycle/visited filters. If the filter is reachable then we choose a next operator to work. If no operator left then we choose other component.

Partial Unfolding. We try to build *B* and *C* if it is required, using notion of sufficient condition of static and dynamic permutability. If the last operators satisfy the sufficient condition of static (dynamic) permutability then *B=0*. Starting delayed to build *C*. In general, this problem is formulated as follows. It is given: current state of environment *E*, operator *a* one of the parallel processes u_1 (other processes are delayed to use some operators in this process, including the process by which it was taken the operator *a*), and delayed parallel processes. In the set of parallel processes it is needed to find the operator *b*, which does not satisfy the sufficient condition of permutability (*table D*) or does not satisfy the condition of sufficient dynamic permutability regarding the *E*. In order to check whether these operators are in a given process, you generally build all states space. If these operators are not found then all resulting state of the search should be removed from storage cycle/visited filters. This is necessary because subtrace which leads to the required operator can modify the current state of the environment and a sufficient condition for dynamic permutability can not be performed, although for state *E* a sufficient condition for dynamic permutability is performed. But in some cases the search for these operators do not need to spend a dynamically performance of all subtrace. If the state of the

environment does not contain predicates that combine memory of these operators and the set of attributes operators intersect, then we can use the concept of specialization[16] in order to break into several operators and sufficient condition will check only those suboperator memory that belongs to the operator. If such operator was found then we add result of it insertion into the list of components.

For trace equivalence each trace for deadlock should be checked additionally because of used normal form. Each cycle/visited trace should be checked for reachability of user defined state in the following way: turning back with a help of backward predicate transformer until operator doesn't have value 1 as first and second parameters in the table *Flt*.

Theorem 6. $punfold(E,u,i)$ Function which was represented in fig. 1 saves property of reachability checking.

Proof.

Let's suppose opposite that the function $punfold(E,u,i)$ does not save the property of reachability checking. This means that for some state of environment E such operator a exists, which is applied to the $E(E \xrightarrow{a} E')$ and doesn't exist as first action in components $A(i), B(E,i), C(E,i)$. So, the operator will be dynamically permutable regarding the environment E and all other operators resulting behavior components $A(i), B(E,i), C(E,i)$ (according to *Choosing Operator, Partial Unfolding*). In addition, operator a can be applied after the application of the first operators in the resulting behavior of components $A(i), B(E,i), C(E,i)$.

This means that the required state of the environment is reachable, but after applying the operator a on the next step these operators are dynamically permutable regarding the environment E. From other point of view we do not take into account E'. If value of pair for the operator in the table *Flt* is (0,1) then according to step 5 we have to take it into consideration and in this case it will be the first operator in the behavior of components $A(i), B(E,i), C(E,i)$. If the value of such pair is (1,1) then the state of the environment is reachable in the next step, as defined *Flt*. If the value is (1,0) then before application of the operator a we need to check the reachability of the environment E and definitions in *Flt.* If the value is (0,0) then required state is not reached in E'. That means that the required state of the environment is not unreached at all states of the environment as a result of unfolding application $punfold(E,u,i)$. So we got contradiction. The theorem is proved.

The main problem of proposed algorithm is complicity to find component $C(E,i)$. One of the ways to speed up such algorithm is delayed computation. The idea of this method contains the following:

1) To collect all such special states from *Partial Unfolding*, where we should find component $C(E,i)$ and finding the required states with a help of different methods: all states coverage, invariants etc.

2) To continue algorithm with built states of component $C(E,i)$.

Such algorithm is called *incremental algorithm of reachability checking*.

6.4. The Static Interleaving Reduction Algorithm

If operators and initial environment state of model do not contain predicates which connected to the memory of different parallel processes then general algorithm in previous section could be simplified. Such algorithm is called *static interleaving reduction algorithm*.

For the component $C(E,i)$ of $punfold(E,u,i)$ we should check reachability of application of operator which is not dynamically permutable for a_i (see section 5). For elimination of such reachability checking we could build additional interleaving according to checking of reachability of corresponded operator in behavior. For example, let $a\,||\,b.c\,||\,d$ and $\neg(a \leftrightarrow c) \wedge \neg(b \leftrightarrow d)$, E be some environment state. So,

$$punfold(E, a\,||\,b.c\,||\,d, 1) = E[a.(b.c)\,||\,d] + E[b.c.(a\,||\,d)] + E[d.(a\,||\,b.c)]$$

Here we take into account $b.c.(a\,||\,d)$, because $\neg(a \leftrightarrow c)$; $d.(a\,||\,b.c)$, because we have taken $b.c.(a\,||\,d)$ and $\neg(b \leftrightarrow d)$.

6.5. Examples of Application

In Table 1 information about few big examples run with our static interleaving reduction algorithm are presented. All of them give out of memory error (PC with 8 Gb of RAM) if we try to obtain all states space. So, we try to run them on implementation of proposed algorithm in Insertion Modeling System.

Table 1. Experiments result for static interleaving reduction algorithm

No.	Total number of operators pairs	Number of non-permutable operator's pairs	Time
1	660	30	25 min (on prototype)
2	780	14	47 min 4 sec
3	12882	225	1 min 42 sec

Here "on prototype" means that this experiment was done on the algorithm which was implemented in language of Insertion Modeling System[17]. Other experiments were run on the algorithm which was implemented on C++. "Total number of operators pairs" is number of pairs of operators which were detected in parallel behavior of the model. "Number of non-permutable operator's pairs" is a number of detected non-permutable pairs of operators. Example 2 works slower because it has four parallel processes and each sequential process more non-deterministic, example 1 has only 2 parallel processes.

7 Conclusion

Described algorithm of interleaving reduction was implemented in Insertional Modeling System. Its restriction for usage of static permutability condition was good account in set of big examples. In any case, the main interleaving reduction algorithm depends on reachability checking problem (component $C(E,i)$, section 6).

Notoriously this problem is algorithmically unsolvable. It means that you could always prepare example where interleaving reduction will be impossible (for example, all operators will be non-permutable etc).

References

1. The Interleaving Paradigm, http://www-i2.informatik.rwth-aachen.de/i2/fileadmin/user_upload/documents/MC08/mc_lec3.pdf
2. Symbolic Modeling, http://en.wikipedia.org/wiki/Model_checking
3. Alessio Lomuscio, Wojciech Penczek, and Hongyang Qu. 2010. Partial Order Reductions for Model Checking Temporal-epistemic Logics over Interleaved Multi-agent Systems. Fundam. Inf. 101, 71-90, 1-2 (January 2010).
4. C. Norris Ip and David L. Dill. 1996. Better Verification through Symmetry. Form. Methods Syst. Des. 9, 41-75, 1-2 (August 1996).
5. Edmund M. Clarke, Orna Grumberg, and David E. Long. Model Checking and Abstraction. *ACM Trans. Program. Lang. Syst.* 16,1512-1542, 5 (September 1994)
6. Vijay D'Silva, Mitra Purandare, and Daniel Kroening. Approximation Refinement for Interpolation-Based Model Checking. In Proceedings of the 9th international conference on Verification, model checking, and abstract interpretation (VMCAI'08), Francesco Logozzo, Doron A. Peled, and Lenore D. Zuck (Eds.), Berlin, Heidelberg, pp. 68-82, Springer-Verlag (2008)
7. Data-Flow Analysis, http://en.wikipedia.org/wiki/Data-flow_analysis.
8. K.L. McMillan: Trace Theoretic Verification of Asynchronous Circuits Using Unfoldings. Proceedings of the 7th Workshop on Computer Aided Verification, Liege, LNCS 939, pp. 180-195, Springer (1995)
9. E. W. Dijkstra. Hierarchical Ordering of Sequential Processes, Acta Informatica 1(2), 115-138. (1971)
10. A. Letichevsky, A. Godlevsky, A. Letichevsky Jr., S. Potienko, V. Peschanenko. Properties of Predicate Transformer of VRS System. Cybernetics and System Analyses 4, 13-16. (2010)
11. Escobar, J. Meseguer: Symbolic Model Checking of Infinite-State Systems Using Narrowing. Proceedings of the 18th International Conference on Term Rewriting and Applications, LNCS 4533, 153-168, Springer (2007).
12. Frédéric Herbreteau, Grrégoire Sutre, and The Quang Tran. 2007. Unfolding Concurrent Well-Structured Transition Systems. In Proceedings of the 13th international conference on Tools and algorithms for the construction and analysis of systems (TACAS'07), Orna Grumberg and Michael Huth (Eds.), Berlin, Heidelberg, 706-720, Springer-Verlag (2007).
13. A. Letichevsky, O. Letychevskyi, V. Peschanenko. About One Efficient Algorithm for Reachability Checking in Modeling and Its Implementation. ICTERI 2012, Communications in Computer and Information Science 149, 149-165. (Springer, 2012)
14. A. Letichevsky, O. Letychevskyi, V. Peschanenko. Insertion Modeling System. PSI 2011, Lecture Notes in Computer Science 7162, 262-274. (Springer, 2011)
15. C. Norris Ip and David L. Dill. 1996. Better Verification through Symmetry. Form. Methods Syst. Des. 9, 41-75, 1-2 (August 1996)
16. V. Peschanenko, A. Guba, C. Shushpanov. Specializations in Symbolic Verification. Communications in Computer and Information Science 412, 332–354, Springer (2013)
17. APS and IMS systems, http://apsystems.org.ua

Abstracting an operational semantics to finite automata

Nadezhda Baklanova, Wilmer Ricciotti, Jan-Georg Smaus, Martin Strecker

IRIT (Institut de Recherche en Informatique de Toulouse)
Université de Toulouse, France
firstname.lastname@irit.fr *,**

Abstract. There is an apparent similarity between the descriptions of small-step operational semantics of imperative programs and the semantics of finite automata, so defining an abstraction mapping from semantics to automata and proving a simulation property seems to be easy. This paper aims at identifying the reasons why simple proofs break, among them artifacts in the semantics that lead to stuttering steps in the simulation. We then present a semantics based on the zipper data structure, with a direct interpretation of evaluation as navigation in the syntax tree. The abstraction function is then defined by equivalence class construction.

Keywords: Programming language semantics; Abstraction; Finite Automata; Formal Methods; Verification
Key Terms: FormalMethod, VerificationProcess

1 Introduction

Among the formalisms employed to describe the semantics of transition systems, two particularly popular choices are abstract machines and structural operational semantics (SOS). Abstract machines are widely used for modeling and verifying dynamic systems, e.g. finite automata, Büchi automata or timed automata [9,4,1]. An abstract machine can be represented as a directed graph with transition semantics between nodes. The transition semantics is defined by moving a pointer to a current node. Automata are a popular tool for modeling dynamic systems due to the simplicity of the verification of automata systems, which can be carried out in a fully automated way, something that is not generally possible for Turing-complete systems.

This kind of semantics is often extended by adding a background state composed of a set of variables with their values: this is the case of timed automata, which use background clock variables [2]. The Uppaal model checker for timed

* N. Baklanova and M. Strecker were partially supported by the project *Verisync* (ANR-10-BLAN-0310).
** W. Ricciotti and J.-G. Smaus are supported by the project Ajitprop of the Fondation Airbus.

automata extends the notion of background state even further by adding integer and Boolean variables to the state [7] which, however, do not increase the computational power of such timed automata but make them more convenient to use.

Another formalism for modeling transition systems is structural semantics ("small-step", contrary to "big-step" semantics which is much easier to handle but which is inappropriate for a concurrent setting), which uses a set of reduction rules for simplifying a program expression. It has been described in detail in [14] and used, for example, for the Jinja project developing a formal model of the Java language [10]. An appropriate semantic rule for reduction is selected based on the expression pattern and on values of some variables in a state. As a result of reduction the expression and the state are updated.

$$\frac{s' = s(v \longmapsto eval\ expr\ s)}{(Assign\ v\ expr, s) \to (Unit, s')}\ [\textsc{Assignment}]$$

This kind of rules is intuitive; however, the proofs involving them require induction over the expression structure. A different approach to writing a structural semantics was described in [3,12] for the CMinor language. It uses a notion of continuation which represents an expression as a control stack and deals with separate parts of the control stack consecutively.

$$(Seq\ e1\ e2 \cdot \kappa, s) \to (e1 \cdot e2 \cdot \kappa, s) \qquad (Empty \cdot \kappa, s) \to (\kappa, s)$$

Here the "$\cdot$" operator designates concatenation of control stacks. The semantics of continuations does not need induction over the expression, something which makes proof easier; however it requires more auxiliary steps for maintaining the control stack which do not have direct correspondance in the modeled language.

For modeling non-local transfer of control, Krebbers and Wiedijk [11] present a semantics using (non-recursive) "statement contexts". These are combined with the above-mentioned continuation stacks. The resulting semantics is situated mid-way between [3] and the semantics proposed below.

The present paper describes an approach to translation from structural operational semantics to finite automata extended with background state. All the considered automata are an extension of Büchi automata with background state, i.e. they have a finite number of nodes and edges but can produce an infinite trace. The reason of our interest in abstracting from structural semantics to Büchi automata is our work in progress [6]. We are working on a static analysis algorithm for finding possible resource sharing conflicts in multithreaded Java programs. For this purpose we annotate Java programs with timing information and then translate them to a network of timed automata which is later model checked. The whole translation is formally verified. One of the steps of the translation procedure includes switching from structural operational semantics of a Java-like language to automata semantics. During this step we discovered some problems which we will describe in the next section. The solutions we propose

extend well beyond the problem of abstracting a structured language to an automaton. It can also be used for compiler verification, which usually is cluttered up with arithmetic adress calculation that can be avoided in our approach.

The contents of the paper has been entirely formalized in the Isabelle proof assistant [13]. We have not insisted on any Isabelle-specific features, therefore this formalization can be rewritten using other proof assistants. The full Isabelle formal development can be found on the web [5].

2 Problem Statement

We have identified the following as the main problems when trying to prove the correctness of the translation between a programming language semantics and its abstraction to automata:

1. Preservation of execution context: an abstract machine always sees all the available nodes while a reduced expression loses the information about previous reductions.
2. Semantic artifacts: some reduction rules are necessary for the functionality of the semantics, but may be missing in the modeled language. Additionally, the rules can produce expressions which do not occur in the original language.

These problems occur independently of variations in the presentation of semantic rules [14] adopted in the literature, such as [10] (recursive evaluation of sub-statements) or [3,12] (continuation-style).

We will describe these two problems in detail, and later our approach to their solution, in the context of a minimalistic programming language which only manipulates Boolean values (a *Null* value is also added to account for errors):

datatype *val* = *Bool bool* | *Null*

The language can be extended in a rather straightforward way to more complex expressions. In this language, expressions are either values or variables:

datatype *expr* = *Val val* | *Var vname*

The statements are those of a small imperative language:

datatype *stmt* =
  Empty — no-op
| *Assign vname val* — assignment: $var := val$
| *Seq stmt stmt* — sequence: $c_1; c_2$
| *Cond expr stmt stmt* — conditional: if e then c_1 else c_2
| *While expr stmt* — loop: while e do c

2.1 Preservation of execution context

Problem 1 concerns the loss of an execution context through expression reductions which is a design feature of structural semantics. Let us consider a simple example.

Assume we have a structural semantics for our minimal imperative language (some rules of a traditional presentation are shown in Figure 1): we want to translate a program written in this language into an abstract machine. Assume that the states of variable values have the same representation in the two systems: this means we only need to translate the program expression into a directed graph with different nodes corresponding to different expressions obtained by reductions of the initial program expression.

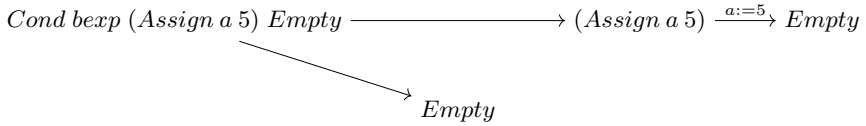
$$\frac{s' = s(v \longmapsto eval\ expr\ s)}{(Assign\ v\ expr, s) \to (Empty, s')}\ [\textsc{Assign}]$$

$$\frac{eval\ bexp\ s = True}{(Cond\ bexp\ e1\ e2, s) \to (e1, s)}\ [\textsc{CondT}] \qquad \frac{eval\ bexp\ s = False}{(Cond\ bexp\ e1\ e2, s) \to (e2, s)}\ [\textsc{CondF}]$$

Fig. 1. Semantic rules for the minimal imperative language.

On the abstract machine level the *Assign* statements would be represented as two-state automata, and the *Cond* as a node with two outgoing edges directed to the automata for the bodies of its branches.

Consider a small program in this language $Cond\ bexp\ (Assign\ a\ 5)\ Empty$ and its execution flow.



The execution can select any of the two branches depending on the *bexp* value. There are two different *Empty* expressions appearing as results of two different reductions. The corresponding abstract machine would be a natural graph representation for a condition statement with two branches (Figure 2).

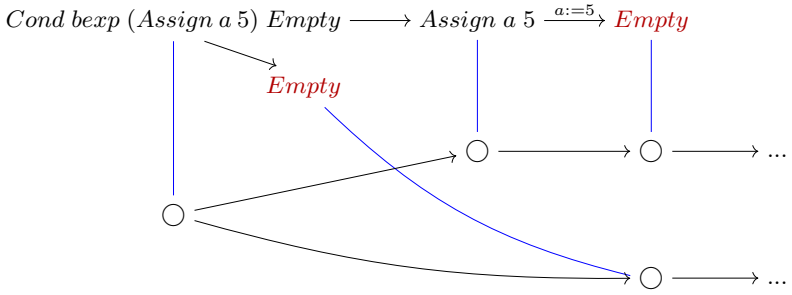


Fig. 2. The execution flow and the corresponding abstract machine for the program $Cond\ bexp\ (Assign\ a\ 5)\ Empty$.

During the simple generation of an abstract machine from a program expression the two *Empty* statements cannot be distinguished although they should be mapped into two different nodes in the graph. We need to add more information about the context into the translation, and it can be done by different ways.

A straightforward solution would be to add some information in order to distinguish between the two *Empty* expressions. If we add unique identifiers to each subexpression of the program, they will allow to know exactly which subexpression we are translating (Figure 3). The advantage of this approach is its simplicity, however, it requires additional functions and proofs for identifier management.

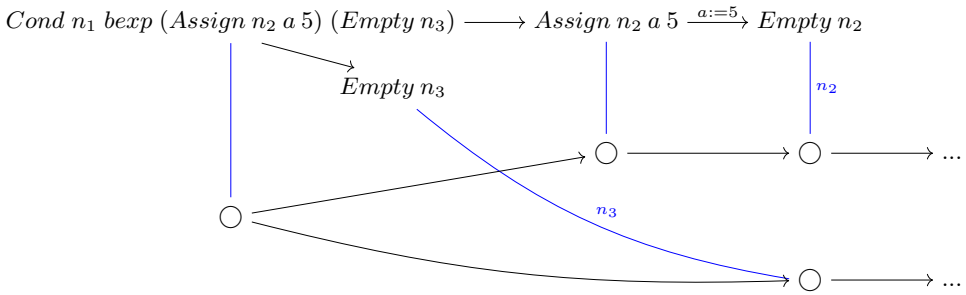


Fig. 3. The execution flow and the corresponding abstract machine for the program with subexpression identifiers *Cond* n_1 *bexp* (*Assign* n_2 *a* 5) (*Empty* n_3).

Another solution for the problem proposed in this paper involves usage of a special data structure to keep the context of the translation. There are known examples of translations from subexpression-based semantics [10] and continuation-based semantics [12] to abstract machines. However, all these translations do not address the problem of context preservation during the translation.

2.2 Semantic artifacts

The second problem appears because of the double functionality of the *Empty* expression: it is used to define an empty operator which does nothing as well as the final expression for reductions which cannot be further reduced. The typical semantic rules for a sequence of expressions look as shown on Figure 4.

$$\frac{(e1, s) \to (e1', s')}{(Seq\ e1\ e2, s) \to (Seq\ e1'\ e2, s')}\ [\textsc{Seq1}] \qquad \frac{}{(Seq\ Empty\ e2, s) \to (e2, s)}\ [\textsc{Seq2}]$$

Fig. 4. Semantic rules for the sequence of two expressions.

Here the *Empty* expression means that the first expression in the sequence has been reduced up to the end, and we can start reducing the second expression. However, any imperative language translated to an assembly language would not have an additional operator between the two pieces of code corresponding to the first and the second expressions. The rule SEQ2 must be marked as a silent transition when translated to an automaton, or the semantic rules have to be changed.

3 Zipper-based semantics of imperative programs

3.1 The zipper data structure

Our plan is to propose an alternative technique to formalize operational semantics that will make it easier to preserve the execution context during the translation to an automata-based formalism. Our technique is built around a zipper data structure, whose purpose is to identify a location in a tree (in our case: a *stmt*) by the subtree below the location and the rest of the tree (in our case: of type *stmt-path*). In order to allow for an easy navigation, the rest of the tree is turned inside-out so that it is possible to reach the root of the tree by following the backwards pointers. The following definition is a straightforward adaptation of the zipper for binary trees discussed in [8] to the *stmt* data type:

```
datatype stmt-path =
  PTop
| PSeqLeft stmt-path stmt          | PSeqRight stmt stmt-path
| PCondLeft expr stmt-path stmt    | PCondRight expr stmt stmt-path
| PWhile expr stmt-path
```

Here, *PTop* represents the root of the original tree, and for each constructor of *stmt* and each of its sub-*stmt*s, there is a "hole" of type *stmt-path* where a subtree can be fitted in. A location in a tree is then a combination of a *stmt* and a *stmt-path*:

```
datatype stmt-location = Loc stmt stmt-path
```

Given a location in a tree, the function *reconstruct* reconstructs the original tree *reconstruct* :: *stmt* ⇒ *stmt-path* ⇒ *stmt*, and *reconstruct-loc* (*Loc c sp*) = *reconstruct c sp* does the same for a location.

```
fun reconstruct :: stmt ⇒ stmt-path ⇒ stmt where
  reconstruct c PTop = c
| reconstruct c (PSeqLeft sp c2) = reconstruct (Seq c c2) sp
| reconstruct c (PSeqRight c1 sp) = reconstruct (Seq c1 c) sp
| reconstruct c (PCondLeft e sp c2) = reconstruct (Cond e c c2) sp
| reconstruct c (PCondRight e c1 sp) = reconstruct (Cond e c1 c) sp
| reconstruct c (PWhile e sp) = reconstruct (While e c) sp


fun reconstruct-loc :: stmt-location ⇒ stmt where
  reconstruct-loc (Loc c sp) = reconstruct c sp
```

3.2 Semantics

Our semantics is a small-step operational semantics describing the effect of the execution a program on a certain program state. For each variable, the state yields *Some* value associated with the variable, or *None* if the variable is unassigned. More formally, the state is a mapping *vname* ⇒ *val option*. Defining the evaluation of an expression in a state is then standard.

Before commenting the rules of our semantics, let us discuss which kind of structure we are manipulating. The semantics essentially consists in moving around a pointer within the syntax tree. As explained in Section 3.1, a position in the syntax tree is given by a *stmt-location*. However, during the traversal of the syntax tree, we visit each position at least twice (and possibly several times, for example in a loop): before executing the corresponding statement, and after finishing the execution. We therefore add a Boolean flag, where *True* is a marker for "before" and *False* for "after" execution.

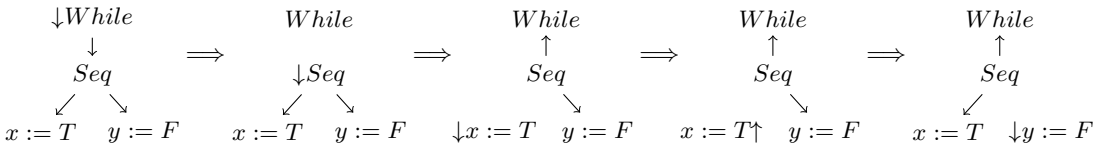


Fig. 5. Example of execution of small-step semantics

As an example, consider the execution sequence depicted in Figure 5 (with assignments written in a more readable concrete syntax), consisting of the initial steps of the execution of the program *While* (*e*, *Seq*(*x* := *T*, *y* := *F*)). The before (resp. after) marker is indicated by a downward arrow before (resp. an upward arrow behind) the current statement. The condition of the loop is omitted because it is irrelevant here. The middle configuration would be coded as ((*Loc* (*x* := *T*) (*PSeqLeft* (*PWhile e PTop*) (*y* := *F*))), *True*).

Altogether, we obtain a syntactic configuration (*synt-config*) which combines the location and the Boolean flag. The semantic configuration (*sem-config*) manipulated by the semantics adjoins the *state*, as defined previously.

type-synonym *synt-config* = *stmt-location* × *bool*
type-synonym *sem-config* = *synt-config* × *state*

The rules of the small-step semantics of Figure 7 fall into two categories: before execution of a statement *s* (of the form ((*l*, *True*), *s*)) and after execution (of the form ((*l*, *False*), *s*)); there is only one rule of this latter kind: SFalse.

Let us comment on the rules in detail:

- SEmpty executes the *Empty* statement just by swapping the Boolean flag.
- SAssign is similar, but it also updates the state for the assigned variable.

```
fun next-loc :: stmt ⇒ stmt-path ⇒ (stmt-location × bool) where
  next-loc c PTop = (Loc c PTop, False)
| next-loc c (PSeqLeft sp c₂) = (Loc c₂ (PSeqRight c sp), True)
| next-loc c (PSeqRight c₁ sp) = (Loc (Seq c₁ c) sp, False)
| next-loc c (PCondLeft e sp c₂) = (Loc (Cond e c c₂) sp, False)
| next-loc c (PCondRight e c₁ sp) = (Loc (Cond e c₁ c) sp, False)
| next-loc c (PWhile e sp) = (Loc (While e c) sp, True)
```

Fig. 6. *Finding the next location*

$$\frac{}{((Loc\ Empty\ sp,\ True),\ s) \rightarrow ((Loc\ Empty\ sp,\ False),\ s)}\ [\textsc{SEmpty}]$$

$$\frac{}{((Loc\ (Assign\ vr\ vl)\ sp,\ True),\ s) \rightarrow ((Loc\ (Assign\ vr\ vl)\ sp,\ False),\ s(vr \mapsto vl))}[\textsc{SAssign}]$$

$$\frac{}{((Loc\ (Seq\ c_1\ c_2)\ sp,\ True),\ s) \rightarrow ((Loc\ c_1\ (PSeqLeft\ sp\ c_2),\ True),\ s)}\ [\textsc{SSeq}]$$

$$\frac{eval\ e\ s = Bool\ True}{((Loc\ (Cond\ e\ c_1\ c_2)\ sp,\ True),\ s) \rightarrow ((Loc\ c_1\ (PCondLeft\ e\ sp\ c_2),\ True),\ s)}[\textsc{SCondT}]$$

$$\frac{eval\ e\ s = Bool\ False}{((Loc\ (Cond\ e\ c_1\ c_2)\ sp,\ True),\ s) \rightarrow ((Loc\ c_2\ (PCondRight\ e\ c_1\ sp),\ True),\ s)}[\textsc{SCondF}]$$

$$\frac{eval\ e\ s = Bool\ True}{((Loc\ (While\ e\ c)\ sp,\ True),\ s) \rightarrow ((Loc\ c\ (PWhile\ e\ sp),\ True),\ s)}\ [\textsc{SWhileT}]$$

$$\frac{eval\ e\ s = Bool\ False}{((Loc\ (While\ e\ c)\ sp,\ True),\ s) \rightarrow ((Loc\ (While\ e\ c)\ sp,\ False),\ s)}\ [\textsc{SWhileF}]$$

$$\frac{sp \neq PTop}{((Loc\ c\ sp,\ False),\ s) \rightarrow (next\text{-}loc\ c\ sp,\ s)}\ [\textsc{SFalse}]$$

Fig. 7. *Small-step operational semantics*

- SSeq moves the pointer to the substatement c_1, pushing the substatement c_2 as continuation to the statement path.
- SCondT and SCondF move to the *then*- respectively *else*- branch of the conditional, depending on the value of the condition.
- SWhileT moves to the body of the loop.
- SWhileF declares the execution of the loop as terminated, by setting the Boolean flag to *False*.

– SFalse comes into play when execution of the current statement is finished. We then move to the next location, provided we have not already reached the root of the syntax tree and the whole program terminates.

The move to the next relevant location is accomplished by function *next-loc* (Figure 6) which intuitively works as follows: upon conclusion of the first substatement in a sequence, we move to the second substatement. When finishing the body of a loop, we move back to the beginning of the loop. In all other cases, we move up the syntax tree, waiting for rule SFalse to relaunch the function.

4 Target language: Automata

4.1 Syntax

As usual, our automata are a collection of nodes and edges, with a distinguished initial state. In this general definition, we will keep the node type $'n$ abstract. It will later be instantiated to *synt-config*. An edge connects two nodes; moving along an edge may trigger an assignment to a variable (*AssAct*), or have no effect at all (*NoAct*).

An automaton $'n\ ta$ is a record consisting of a set of *nodes*, a set of *edges* and an initial node *init-s*. An edge has a *source* node, an *action* and a destination node *dest*. Components of a record are written between ⦇ ... ⦈.

4.2 Semantics

An automaton state is a node, together with a *state* as in Section 3.2.

type-synonym $'n\ ta\text{-}state = 'n * state$

Executing a step of an automaton in an automaton state $(l,\ s)$ consists of selecting an edge starting in node l, moving to the target of the edge and executing its action. Automata are non-deterministic; in this simplified model, we have no guards for selecting edges.

$$\frac{e \in set\ (edges\ aut) \qquad l = source\ e \qquad l' = dest\ e \qquad s' = action\text{-}effect\ (action\ e)\ s}{aut \vdash (l,\ s) \rightarrow (l',\ s')} \text{ [Action]}$$

5 Automata construction

The principle of abstracting a statement to an automaton is simple; the novelty resides in the way the automaton is generated via the zipper structure: as nodes, we choose the locations of the statements (with their Boolean flags), and as edges all possible transitions of the semantics.

To make this precise, we need some auxiliary functions. We first define a function *all-locations* of type *stmt* ⇒ *stmt-path* ⇒ *stmt-location list* which gathers all locations in a statement, and a function *nodes-of-stmt-locations* which adds the Boolean flags.

As for the edges, the function *synt-step-image* yields all possible successor configurations for a given syntactic configuration. This is of course an over-approximation of the behavior of the semantics, since some of the source tree locations may be unreachable during execution.

```
fun synt-step-image :: synt-config ⇒ synt-config list where
  synt-step-image (Loc Empty sp, True) = [(Loc Empty sp, False)]
| synt-step-image (Loc (Assign vr vl) sp, True) = [(Loc (Assign vr vl) sp, False)]
| synt-step-image (Loc (Seq c1 c2) sp, True) = [(Loc c1 (PSeqLeft sp c2), True)]
| synt-step-image (Loc (Cond e c1 c2) sp, True) =
          [(Loc c1 (PCondLeft e sp c2), True), (Loc c2 (PCondRight e c1 sp), True)]
| synt-step-image (Loc (While e c) sp, True) =
           [(Loc c (PWhile e sp), True), (Loc (While e c) sp, False)]
| synt-step-image (Loc c sp, False) = (if sp = PTop then [] else [next-loc c sp])
```

Together with the following definitions:

```
fun action-of-synt-config :: synt-config ⇒ action where
  action-of-synt-config (Loc (Assign vn vl) sp, True) = AssAct vn vl
| action-of-synt-config (Loc c sp, b) = NoAct

definition edge-of-synt-config :: synt-config ⇒ synt-config edge list where
edge-of-synt-config s =
map(λ t. (|source = s, action = action-of-synt-config s, dest = t|))(synt-step-image s)
definition edges-of-nodes :: synt-config list ⇒ synt-config edge list where
  edges-of-nodes nds = concat (map edge-of-synt-config nds)
```

we can define the translation function from statements to automata:

```
fun stmt-to-ta :: stmt ⇒ synt-config ta where
  stmt-to-ta c =
  (let nds = nodes-of-stmt-locations (all-locations c PTop) in
   (| nodes = nds, edges = edges-of-nodes nds, init-s = ((Loc c PTop), True) |))
```

6 Simulation Property

We recall that the nodes of the automaton generated by *stmt-to-ta* are labeled by configurations (location, Boolean flag) of the syntax tree. The simulation lemma (Lemma 1) holds for automata with appropriate closure properties: a successor configuration wrt. a transition of the semantics is also a label of the automaton (*nodes-closed*), and analogously for edges (*edges-closed*) or both nodes and edges (*synt-step-image-closed*).

The simulation statement is a typical commuting-diagram property: a step of the program semantics can be simulated by a step of the automaton semantics,

for corresponding program and automata states. For this correspondence, we use the notation $\approx$, even though it is just plain syntactic equality in our case.

Lemma 1 (Simulation property).
Assume that synt-step-image-closed aut and $(((lc,\ b),\ s) \approx ((lca,\ ba),\ sa))$. *If* $((lc,\ b),\ s) \rightarrow ((lc',\ b'),\ s')$, *then there exist* lca', ba', sa' *such that* $(lca',\ ba') \in$ *set* (*nodes aut*) *and the automaton performs the same transition:* $aut \vdash ((lca,\ ba),\ sa) \rightarrow ((lca',\ ba'),\ sa')$ *and* $((lc',\ b'),\ s') \approx ((lca',\ ba'),\ sa')$.

The proof is a simple induction over the transition relation of the program semantics and is almost fully automatic in the Isabelle proof assistant.

We now want to get rid of the precondition *synt-step-image-closed aut* in Lemma 1. The first subcase (edge closure), is easy to prove. Node closure is more difficult and requires the following key lemma:

Lemma 2.
If $lc \in$ *set* (*all-locations c PTop*) *then set* (*map fst* (*synt-step-image* $(lc,\ b)$)) $\subseteq$ *set* (*all-locations c PTop*).

With this, we obtain the desired

Lemma 3 (Closure of automaton). *synt-step-image-closed* (*stmt-to-ta c*)

For the proofs, see [5].

Let us combine the previous results and write them more succinctly, by using the notation $\rightarrow^*$ for the reflexive-transitive closure for the transition relations of the small-step semantics and the automaton. Whenever a state is reachable by executing a program c in its initial configuration, then a corresponding $(\approx)$ state is reachable by running the automaton generated with function *stmt-to-ta*:

Theorem 1.
If $((Loc\ c\ PTop,\ True),\ s) \rightarrow^* (cf',\ s')$ *then* $\exists\, cfa'\ sa'.$ *stmt-to-ta* $c \vdash$ (*init-s* (*stmt-to-ta c*), s) $\rightarrow^* (cfa',\ sa') \ \wedge\ (cf',\ s') \approx (cfa',\ sa')$.

Obviously, the initial configuration of the semantics and the automaton are in the simulation relation $\approx$, and for the inductive step, we use Lemma 1.

7 Conclusions

This paper has presented a new kind of small-step semantics for imperative programming languages, based on the zipper data structure. Our primary aim is to show that this semantics has decisive advantages for abstracting programming language semantics to automata. Even if the generated automata have a great number of silent transitions, these can be removed.

We are currently in the process of adopting this semantics in a larger formalization from Java to Timed Automata [6]. As most constructs (zipper data

structure, mapping to automata) are generic, we think that this kind of semantics could prove useful for similar formalizations with other source languages. The proofs (here carried out with the Isabelle proof assistant) have a pleasingly high degree of automation that are in sharp contrast with the index calculations that are usually required when naming automata states with numbers.

Renaming nodes from source tree locations to numbers is nevertheless easy to carry out, see the code snippet provided on the web page [5] of this paper. For these reasons, we think that the underlying ideas could also be useful in the context of compiler verification, when converting a structured source program to a flow graph with basic blocs, but before committing to numeric values of jump targets.

References

1. Rajeev Alur, Costas Courcoubetis, and David L. Dill. Model-checking for real-time systems. In *LICS*, pages 414–425. IEEE Computer Society, 1990.
2. Rajeev Alur and David L. Dill. A theory of timed automata. *Theoretical Computer Science*, 126:183–235, 1994.
3. Andrew W. Appel and Sandrine Blazy. Separation logic for small-step cminor. In *Theorem Proving in Higher Order Logics, 20th int. conf. TPHOLS*, pages 5–21. Springer, 2007.
4. Ch. Baier and J.-P. Katoen. *Principles of Model Checking*. MIT Press, 2008.
5. Nadezhda Baklanova, Wilmer Ricciotti, Jan-Georg Smaus, and Martin Strecker. Abstracting an operational semantics to finite automata (formalization), 2014. https://bitbucket.org/Martin_Strecker/abstracting_op_sem_to_automata.
6. Nadezhda Baklanova and Martin Strecker. Abstraction and verification of properties of a Real-Time Java. In *Proc. ICTERI*, volume 347 of *Communications in Computer and Information Science*, pages 1–18. Springer, 2013.
7. Johan Bengtsson and Wang Yi. Timed automata: Semantics, algorithms and tools. In *Lectures on Concurrency and Petri Nets*, LNCS, pages 87–124. Springer, 2004.
8. Gérard Huet. Functional pearl: The zipper. *Journal of Functional Programming*, 7(5):549–554, September 1997.
9. Bakhadyr Khoussainov and Anil Nerode. *Automata Theory and Its Applications*. Birkhauser Boston, 2001.
10. Gerwin Klein and Tobias Nipkow. A machine-checked model for a Java-like language, virtual machine, and compiler. *ACM Trans. Program. Lang. Syst.*, 28:619–695, July 2006.
11. Robbert Krebbers and Freek Wiedijk. Separation logic for non-local control flow and block scope variables. In Frank Pfenning, editor, *Foundations of Software Science and Computation Structures*, volume 7794 of *Lecture Notes in Computer Science*, pages 257–272. Springer Berlin Heidelberg, 2013.
12. Xavier Leroy. A formally verified compiler back-end. *Journal of Automated Reasoning 43(4).*, 43(4), 2009.
13. Tobias Nipkow, Lawrence Paulson, and Markus Wenzel. *Isabelle/HOL. A Proof Assistant for Higher-Order Logic*, volume 2283 of *LNCS*. Springer, 2002.
14. Glynn Winskel. *The Formal Semantics of Programming Languages: An Introduction*. MIT Press, Cambridge, MA, USA, 1993.

The Static Analysis of Linear Loops

Michael Lvov[1], Yulia Tarasich[1],

[1]Kherson State University, 40 rokiv Zhovtnya St. 27
73000, Kherson, Ukraine
{Lvov, YuTarasich}@ksu.ks.ua

Abstract. In the first part of the paper, we consider the problem of generation of polynomial invariants of iterative loops with operator of initialization of loop and non-singular linear operator in the loop body. In the article we also show the algorithm for calculating the basic invariants for linear operator of the Jordan cell, and an algorithm for calculating the basic invariants of diagonalizable linear operator with an irreducible minimal characteristic polynomial. The second part presents a new method for proving the invariance of the system of linear inequalities and of termination of certain linear iterative loops of imperative programs whose data are elements of the constructive linearly ordered field. The theoretical material of the paper is illustrated by examples.

Keywords. Static program analysis, polynomial invariant of a loop, invariant system of linear inequalities, eigenpolynomial of a linear operator.

Key Terms. VerificationProcess, Method, FormalMethod

1 Introduction

As for now, methods of program statistical analysis are being studied intensely. One of the important problems is a problem of the automatic generation of program invariants. Invariants of program are used particularly in methods of programs verification.

The problem of searching for loop invariants in imperative programs was offered by R. Floyd [1] and C. Hoare [2].

A correctness property of the program is formulated in terms of its total or partial correctness. Often, the proof of termination of the program should be implemented separately from the proof of its partial correctness. The algorithmic unsolvability of the termination problem shows that the general algorithm of proof of termination of the program does not exist. To prove the partial correctness of programs, P. Floyd and S. Hoare offered the idea of building loop invariants [1] and invariant relations in control points of programs [2], which allows to prove programs by method of math induction.

Thus, there is a problem of finding the invariants of the program as a key problem of analysis of programs properties.

Now, the main attention is paid to the problem of constructing polynomial invariant equalities. A set of invariant equalities forms the polynomial ideal, a finite basis of which one must build. Note that in a general case, the problem of constructing this basis has not been solved.

The existence and efficiency of algorithms to generate program invariants depend on the subject domain, i.e., on the properties of the data algebras the program deals with. Problems of automatic generation of program invariants for various data algebras have been being analyzed since beginning of 1970s at the Institute of cybernetics of NAS of Ukraine. Their main results are represented in [3,4].

Numerical data algebras are the most important from the practical point of view. The paper [5] outlines two methods of constructing polynomial invariant equalities types in programs whose data algebra is the domain of integrity (polynomially determinate programs) or a field (rationally determinate programs).

This idea used in [6] to generate polynomial invariants of bounded degree for polynomially determined programs. Program conditions such as $f(X) \neq 0$ were taken into account, where $f(X)$ are polynomials of program variables. In [7] they proposed a method to generate polynomial program invariants of bounded degree in linearly determinate (affine) programs containing recursive procedure calls.

In [8] they proposed a method to generate polynomial loop invariants as template polynomials with the use of the algorithm for computing Grobner bases. In [9] they described a method to generate nonlinear and, generally speaking, nonpolynomial invariant relation for linear loops. The method uses eigenvalues and eigenvectors of the linear operator in the loop body.

The paper [10] is devoted to the algebraic fundamentals of the problem of generating polynomial loop invariants. The main result of the study is an algorithm for generating all polynomial invariants for loops with so-called solvable assignment operators. In particular, affine operators with positive real eigenvalues are solvable. The same authors [11] proposed a method to generate polynomial loop invariants, including enclosed loops, as well as program conditions in the form of both polynomial equalities and inequalities. The paper considers a great number of examples and presents tables for the algorithm time depending on technical parameters of the program being analyzed.

In [12] they proposed an algorithm to search for loop invariants based on a system of recurrent relations with loop variables and parameter *n,* which is the loop index. The algorithm searches for the solution of this system not depended on *n.* It is implemented in Theorema software system and is illustrated with examples in detail.

The problem of the description of invariant inequalities is less studied. The main intricacy lies in the infinity of the basis of the metaideal [13] of polynomial inequalities [13, 14]. Iterative methods for solving the problem of the description of linear invariant inequalities were considered in [15-18]. In [15], the problem of generation of the simplest invariant inequalities is solved. In [16-17], general iterative methods are used to solve the problem of searching for linear invariant inequalities.

In [19] they described a method of proving the invariance of the system of linear inequalities for a class of linear iterative loops with real eigennumbers of linear

operators in the loop body. This method can be applied to the entire class of linear iterative loops and it can also be applied to prove their termination. The paper with description of it is under preparation for a publication.

2 The Static Analysis of Polynomial Invariant Equations

2.1 L-invariants of Linear Maps and Invariants of Linear Loops.

Definition 1. Let W be an *n*-dimensional vector space over the field of rational numbers Q and let $\bar{Q}$ be the algebraic closure of the field Q. Let $X = (x_1,....x_n)$ be an n-dimensional vector of variables. A rational function $p(X) \in \bar{Q}(X)$ is called ***L***-invariant of a linear operator $A : W \to W$ if, for any vector $b \in W$ the following relationship holds:

$$p(A \cdot b) = p(b) \tag{1}$$

Example 1. (a linear operator with characteristic polynomial $x^3 - 2$)
Let us consider a linear operator with the matrix

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 2 & 0 & 0 \end{pmatrix}, X = (x, y, z).$$

It's easy to calculate [26], that the rational expression

$$p(x, y, z) = \frac{(\lambda_1^2 x + \lambda_1 y + z)(\lambda_3^2 x + \lambda_3 y + z)}{(\lambda_2^2 x + \lambda_2 y + z)^2} \tag{2}$$

where $\lambda_1 = \sqrt[3]{2}, \lambda_2 = \sqrt[3]{2}\varepsilon, \lambda_3 = \sqrt[3]{2}\varepsilon^2$, and $\varepsilon = \cos(\frac{2\pi}{3}) + i\sin(\frac{2\pi}{3})$ is the primitive third root of unity, is the L-invariant of this operator.

Definition 2. Let $X = (x_1,..., x_n)$ and $b = (b_1,..., b_n)$ - be two collections of variables. The following fragment of an imperative program is called a linear loop:

```
X := b;
While Q(X, b) do X := A*X
```

Remark 1. Operators X:=b and X:=A*X are interpreted as simultaneous assignments of the values of the variables of the right sides to the variables on the left sides. In what follows, we ignore the condition Q(X, b) and consider that the linear loop is infinite and that its execution is nondeterministic. Thus, we consider loops of the form

```
X := b;
While True|False do X := A*X                    (3)
```

Definition 3. Let a vector $b^{(0)} = (b_1^{(0)},...,b_n^{(0)}) \in W$ be chosen as initial. Sequence of vectors, set by recurrent correlation $b^{(j+1)} = Ab^{(j)}$, will be called the orbit of linear operator A.

A loop sets the orbit of linear operator A in space W. Obviously, an orbit A lies in some one-dimensional variety, and the system of invariants characterizes this variety as algebraic.

Definition 4. Polynomial $P(b, X)$ is called loop invariant if, for any natural j and any $b^{(0)}$ $P(b^{(0)}, b^{(j)}) = 0$.

Theorem 1. If $p(X) = r(X)/q(X)$ is an L-invariant of a linear operator A, then the polynomial $r(X)q(b) - q(X)r(b)$ is an invariant of a linear loop over the field $\bar{Q}$.

We call such loop invariants *L*-invariants (of linear loops).

Example 2. (a linear loop with operator from example 1)

The linear loop corresponding to the operator A, has the form

```
(x, y, z) := (a, b, c);
While True|False do (x, y, z) := (y, z, 2*x)
```

L-invariant of this loop is defined by formula (2):

$$P(x,y,z,a,b,c) = (\lambda_1^2 x + \lambda_1 y + z)(\lambda_3^2 x + \lambda_3 y + z)(\lambda_2^2 a + \lambda_2 b + c)^2 - \\ - (\lambda_2^2 x + \lambda_2 y + z)^2 (\lambda_1^2 a + \lambda_1 b + c)(\lambda_3^2 a + \lambda_3 b + c) \quad (4)$$

Note that L-invariant of the loop $P(x, y, z, a, b, c)$ is defined over a field $\bar{Q}(\lambda_1, \lambda_2, \lambda_3)$. However, it has a set of L-invariants with coefficients from the field Q, which can be constructed, they are shown in (4) the canonical form to the polynomial from $\lambda_1, \lambda_2, \lambda_3$, and then - to the polynomial from λ_2 with using the relation $\lambda_1\lambda_3 = \lambda_2^2$ and Vieta's relation. Technique for computing L-invariants over a field Q is demonstrated in [20]. Note that if the variables a, b, c are the assigned numeric values, L-invariant is converted into a loop invariant.

In [22] they described the results, that link L-invariants to eigenvalues and eigenvectors of the operator A^T. The main result of this work:

Theorem 2 (about the multiplicative relations). Let $\lambda_1,...,\lambda_m$ be eigenvalues of a linear operator A and let $s_1,...,s_m$ be eigenvectors of the conjugate operator A^T

that correspond to these eigenvalues. We assume that there are integers $k_1,...,k_m$ such that

$$\lambda_1^{k_1} \cdot ... \cdot \lambda_m^{k_m} = 1 . \quad (5)$$

Then

$$p(X) = (s_1, X)^{k_1} \cdot ... \cdot (s_m, X)^{k_m} \quad (6)$$

is *L*-invariant of the linear operator A.

Proof of the theorem 2 can be found in [21]

Example 3 (continuation of example 2). Apply the theorem 2 to the example 2. Calculate the eigennumbers of operator A.

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 2 & 0 & 0 \end{pmatrix}, \; h(\lambda) = |A - \lambda E| = \begin{vmatrix} -\lambda & 1 & 0 \\ 0 & -\lambda & 1 \\ 2 & 0 & -\lambda \end{vmatrix} = -\lambda^3 + 2 .$$

A characteristic polynomial has the form b$h(x) = x^3 - 2$. Its roots are $\lambda_1 = \sqrt[3]{2}, \; \lambda_2 = \sqrt[3]{2}\varepsilon, \; \lambda_3 = \sqrt[3]{2}\varepsilon^2$, where $\varepsilon = \exp(i\,2\pi/3)$ is the primitive cube root of unity.

Calculate the eigenvectors s_1, s_2, s_3 of matrix $A^T = \begin{pmatrix} 0 & 0 & 2 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$:

$s_1 = (\lambda_1^2, \lambda_1, 1), \; s_2 = (\lambda_2^2, \lambda_2, 1), \; s_3 = (\lambda_3^2, \lambda_3, 1)$.

It is easy to check that $\frac{\lambda_1 \lambda_3}{\lambda_2^2} = 1$. By the theorem 2 the operator A has a L-invariant (2).

Corollary 1. If the minimum characteristic polynomial $h(x)$ of linear operator A has a free term equal to ± 1 (i.e. $\det(A) = \pm 1$), then the linear operator A has a **L**-invariant.

Example 4. A loop of the points rotation of a plane (a,b) at an angle $\arctan(4/3)$.

```
(x,y) := (a,b);
While True do (x, y):= (4/5*x - 3/5*y, 3/5*x + 4/5*y)
```

Calculate the eigenvalues and eigenvectors of the operator A:

$$A=\begin{pmatrix}4/5 & -3/5\\ 3/5 & 4/5\end{pmatrix}.\ h(\lambda)=|A-\lambda E|=\lambda^2-\frac{8}{5}\lambda+1.$$

$$\lambda_1=\frac{4}{5}-i\frac{3}{5},\lambda_2=\frac{4}{5}+i\frac{3}{5}.\ \ s_1=(i,1),\ s_2=(-i,1).$$

Since $\lambda_1\lambda_2=1$, L-invariant of the operator A is

$$p(x,y)=(ix+y)(-ix+y)=x^2+y^2.$$

And the loop invariant is $x^2+y^2-a^2-b^2$.

Example 5. Loop of Fibonacci sequence calculation, starting with a pair of (a,b).

```
(x,y) := (a,b);
While True|False do (x, y):= (x + y, x)
```

Calculate the eigenvalues and eigenvectors of the operator A:

$$A=\begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix}.\ h(\lambda)=|A-\lambda E|=\lambda^2-\lambda-1.$$

$$\lambda_1=\frac{1}{2}-\frac{1}{2}\sqrt{5},\ \lambda_2=\frac{1}{2}+\frac{1}{2}\sqrt{5}.$$

$$s_1=(\lambda_1,1)=(\frac{1}{2}-\frac{1}{2}\sqrt{5},1),\ s_2=(\lambda_2,1)=(\frac{1}{2}+\frac{1}{2}\sqrt{5},1).$$

Since $\lambda_1\lambda_2=-1$, L-invariant of the operator A is

$$p(x,y)=((\lambda_1x+y)(\lambda_2x+y))^2=(x^2-xy-y^2)^2.$$

The invariant relation of loop is $(x^2-xy-y^2)^2=(a^2-ab-b^2)^2$.

Corollary 2. If the characteristic (minimum) polynomial $h(X)$ of linear operator A is x^m-a, then linear operator has an L-invariants.

Proofs of corollaries 1 and 2 are in [21]

Theorem 3. Let $h(x)$ be an polynomial from variable x with rational coefficients and $\Lambda=(\lambda_1,...,\lambda_m)$ are all its roots in an algebraic closure $\bar{Q}$ of the field Q. Consider the set $G(h)=\{x_1^{k_1}\cdot...\cdot x_m^{k_m}:\lambda_1^{k_1}\cdot...\cdot\lambda_m^{k_m}=1\}$ that is the set of monomials of the field of rational expressions $Q(X)$ (possibly with negative degrees), who receive a value of 1 when we substitute λ_i instead of x_i. Then $G(h)$ is a multiplicative abelian group with a finite number of generators.

The proof of theorem 3 is obvious, since the subgroup of an abelian group with a finite number of generators has a finite number of generators.

It follows from theorem 3 that the main problem for the generation of L-invariants is the problem of finding an algorithm for constructing a set that generate the groups $G(h)$.

Example 6 (continuation of example 3). It is easy to see that we have the following multiplicative relations for the polynomial $h(x) = x^3 - 2$ between its roots:

$$\lambda_1^2 = \lambda_2\lambda_3,\ \lambda_1\lambda_2 = \lambda_3^2,\ \lambda_1\lambda_3 = \lambda_2^2,\ \lambda_2^3 = \lambda_3^3$$

These relations have relevant binomials

$$x_1^2 - x_2x_3,\ x_1x_2 - x_3^2,\ x_1x_3 - x_2^2,\ x_2^3 - x_3^3,$$

that form a Gröbner basis of the ideal $I(G_B) = I(G(h))$.

Corollary 3. The set of all L-invariant of operator A defines the field of rational expressions.

Proof of **corollary** 3 is in [21]

Theorem 4 Let $f(x)$ be irreducible over the field Q and reduced polynomial and $\{\lambda_1, \lambda_2, ...\lambda_m\}$ is the set of its roots over the field $\bar{Q}$. If we have a nontrivial multiplicative relationship $\lambda_1^{k_1}...\lambda_m^{k_m} = 1$ with integer indices $k_1, ..., k_m$ between his roots, then the free term a_m $f(x)$ equal to ± 1 or $\sum_{i=1}^{m} k_i = 0$.

The proof is in [21]

Definition 5. L-invariants of operator A, defined of multiplicative relation between the roots of the characteristic polynomial $\lambda_1 \cdot ... \cdot \lambda_m = \pm 1$, will be called whole. L-invariants of operator A, defined of multiplicative relation $\lambda_1^{k_1} \cdot ... \cdot \lambda_m^{k_m} = 1, \sum k_i = 0$, will be called rational.

Theorem 5. If the characteristic polynomial of operator A is $h(x^k), k > 1$, then operator A has a rational L-invariants.

The **proof** of theorem 5 is in [21]

2.2 L-invariants of Jordan Cells

A nondegenerate linear operator A can be represented in a suitable basis by the following Jordan form of its matrix [18, 22].

$$A = \begin{bmatrix} J_1(\lambda_1) & 0 & ... & 0 \\ 0 & J_2(\lambda_2) & ... & 0 \\ . & . & ... & . \\ 0 & 0 & ... & J_m(\lambda_m) \end{bmatrix}, \qquad (7)$$

where $J_i(\lambda_i)$ are Jordan cells of different sizes. Jordan cell is of the form

$$J(\lambda)=\begin{bmatrix}\lambda & 1 & ... & 0\\ 0 & \lambda & ... & 0\\ 0 & ... & \lambda & 1\\ 0 & ... & 0 & \lambda\end{bmatrix} \qquad (8)$$

Thus, theorem 2 is applied only to the rows of the matrix of the linear operator A, that correspond to the eigenvectors of A, i.e., to the collection of the last rows of Jordan cells $J_i(\lambda_i)$, $i=1,...,m$. Below, we will extend this theorem to arbitrary nondegenerate linear operators by considering Jordan cells on the whole.

Transformation $J := J * X$, where $X=(x_1,...,x_k)$, in the coordinate form is

$$x_1 := \lambda x_1 + x_2; ...; x_{k-1} := \lambda x_{k-1} + x_k; x_k := \lambda x_k$$

Introduce the following notation: $x_{k-1} \overset{df}{=} y,\ x_k \overset{df}{=} z$.

For each Jordan cell $J_k(\lambda_k)$ of the Jordan form of the operator A its own sequence of subspaces of eigenpolynomials is determined.

The main theory of the eigenpolynomials of Jordan cells as well as of the relationship between eigenpolynomials and *L*-invariants of linear operators is formulated in [23, 24].

The concept of eigenpolynomial of a linear operator can be of an independent interest for linear algebra applications.

If all eigennumbers of linear operator A are rational numbers, then the problem of constructing this basis is an algorithmically solvable with the help of theoretical&number algorithm.

In the [25] a direct method of finding invariants of Jordan cells is described. The main results of this work are discussed below.

Theorem 6 (about the structure of the ideal of invariants)**.** Let A be an arbitrary nondegenerate linear operator, presented in a suitable basis of matrix (7), $I_{J_1}(A),...,I_{J_k}(A)$ are ideals of his invariants, presented in homogeneous coordinates

$$u_{ij}=x_{ij}/z_i,\ u_i=y_i/z_i \qquad e_{ij}=a_{ij}/c_i, e_i=b_i/c_i$$

by basis of the form

$$u_j - q_j(\lambda,u,1),\ j=1,...,n-2$$

and $I_\Lambda(A)$ is an ideal of invariants of the operator A_{red}, and $I(A)$ is an ideal of invariants of the operator A (of the loop (3)). Then

$$GBase(I_\Lambda(A)) = GBase(I_{J_1}(A)) \cup ... \cup GBase(I_{J_k}(A)) \cup GBase(I_\Lambda(A))$$

Theorem 7. If a group of multiplicative relations of roots of an irreducible polynomial $f(x)$ is nontrivial $(MR(f)\neq(e))$, there may be two situations:

1. The set of roots $\Lambda = (\lambda_1,...,\lambda_n)$ is divided into certain number l of equally-powerful classes $\Lambda_1,...,\Lambda_l;\ \Lambda_j = \{\lambda_{(j-1)d+1},...\lambda_{jd}\}; j = 1,...,l.$ wherein $d = len(\Lambda_j),\ n = ld$. Multiplicative relations from $MR(f)$ in this situation have the form $\Lambda_j = \varepsilon_j,\ j = 1,...,l$, where ε_j are roots from 1.
2. The equally-powerful classes $\Lambda_1,...,\Lambda_l, \Lambda_i = \{\lambda_{(i-1)d+1},...\lambda_{id}\}; i = 1,...,k.$ Wherein $d = len(\Lambda_j),\ n = kd$. Multiplicative relations from $MR(f)\ MR(f)$ in this situation have the form $\Lambda_i = \varepsilon_{ij}\Lambda_j,\ i = 1,...,l$, where ε_j are roots from 1.

Both situations may occur simultaneously.

For the proof of theorem 7, take a look in [25]. This theorem has a key role for the algorithm of calculation of the system generators of the group $MR(f)$.

Theorem 8. Let $f(x) \in Q[x]$ is an irreducible polynomial and $\lambda_1,...,\lambda_m$ are its roots. The problem of constructing a basis of a set of generating the group $G_U(h) = \{x_1^{k_1}...x_m^{k_m} : \lambda_1^{k_1}...\lambda_m^{k_m} \in U\}$, where U is a group of all roots from 1 is algorithmically solvable.

The proof of theorem 8 is in [25].

Thus, by theorem 6, the invariants of a linear operator can be classified as intracellular - that are inherent to each Jordan cell of linear operator, and intercellular - those that are inherent in its diagonalisable part.

Intracellular invariants are computed directly from the formulas of [25]

$$x_j = \frac{z}{c}(a_j + \frac{C_1(\lambda\frac{cy-bz}{cz})}{\lambda}a_{j+1} + ... + \frac{C_{n-j}(\lambda\frac{cy-bz}{cz})}{\lambda^{n-j}}a_n)\ .$$

The existence of intercellular invariants depend on the existence of nontrivial multiplicative relations between the eigenvalues of the linear operator (theorem 2).

For linear operators with an irreducible minimum characteristic polynomial problem of constructing a basis of set of multiplicative relations between its eigenvalues is algorithmically solvable, but the algorithm of theorem 8 is ineffective due to a very large degree of the polynomial $S(x)$, which is be necessary to decompose into factors.

The problem of constructing a basis of set of multiplicative relations for arbitrary linear operators is still open.

3 The Static Analysis of Linear Inequalities.

Let $W = K^n$ be an n-dimensional vector space over a linearly ordered and constructive field K and $\bar{K}$ is an algebraic closure of K.

Definition 6. As a linear semi-algebraic set $M(x_1,...,x_n)$ is called the area W, that is defined by a quantifier-free formula in the signature of the logical connectives $<\vee,\&,\neg>$ with linear inequalities in the variables $x_1,...,x_n$ as atoms. If the field M is given by the formula $F(X)$, i.e. $M=\{X : F(X)\}$, We shall denote it by $M(F(X))$.

Definition 7. Let $X=(x_1,...,x_n)$, and $\bar{b}=(b_1,...,b_n)$ be two vectors of variables. The linearly loop with the precondition is a fragment of imperative program in the form

```
X := b; //  S(b̄)  - a precondition
While U(X, b) do X := A*X                                    (9)
```

where $S(\bar{b})$, and $U(X,\bar{b})$ are quantifier-free formulas of applied logic of linear semi-algebraic sets, A is a matrix of the linear operator $W \to W$.

Non-deterministic and associated with loop (9) we call the loop of the form

```
X := b; //  S(b̄) - a precondition
While True|False do X := A*X                                 (10)
```

whose number of repeats is nondeterministic.

Remark 2. Definition 7 of loops differs from the definitions 2 and 3 because of its precondition $S(\bar{b})$ that limited the initial values of the loops variables by a linear semi-algebraic set and an introduction to the consideration of the conditions of the loop $U(X,\bar{b})$.

Definition 8. Linear inequality $P(X,b)\in K^1[X,b]$ is called an invariant for the loop (9) with a precondition $S(b)$, if it is executed whenever the loop body is executed.

$$P(X,b)\overset{df}{=}a_1x_1+...+a_nx_n<a_1'b_1+...+a_n'b_n$$

Thus, the invariance means performing of a sequence of formulas

$S(b)\to P(b,b)$, // Invariant is executed in the input in the loop

$U(b,b)\to P(Ab,b)$, // Invariant is executed after the first iteration

$U(Ab,b)\to P(A^2b,b)$, // Invariant is executed after the second iteration

…

$U(A^kb,b)\to P(A^{k+1}b,b)$, // Invariant is executed after the k-th iteration

$\neg U(A^kb,b)\to P(A^kb,b)$ // Invariant is executed at the completion of the loop

Theorem 9. If all eigenvalues $\Lambda = (\lambda_1,...,\lambda_n), \lambda_i \in \overline{K}$ of operator A are real, the problem of proving of the invariance $P(X,b)$ for the loop (9) is algorithmically solvable.

The main content of the proof of theorem 9 is formulated in lemmas 1-5 [13].

Definition 9. The linearly defined loop (10) is called completed if for any $\overline{b} \in M(S(X))$ the sequence

$$\overline{b}^{(0)} = \overline{b},\ \overline{b}^{(m+1)} = A\overline{b}^{(m)}, m = 0,1,... \tag{11}$$

for some natural $m^* = m^*(\overline{b})$ satisfies the relationship $\neg U(\overline{b}^{(m^*)}, \overline{b})$.

Thus, if the loop is completed, for each $\overline{b} \in M(S(X))$ is the smallest positive integer $m^*(\overline{b})$, on which the loop (9) is completed.

Definition 10. Let $\overline{a}, \overline{c} \in K^n$. A linear inequality

$$L(\overline{a}, \overline{c}, X, \overline{b}) \overset{df}{=} (\overline{a}, X) \le (\overline{c}, \overline{b}) \tag{12}$$

is called conditional invariant of linear certain loop (9) (with a precondition $S(\overline{b})$), if for any $\overline{b} \in M(S(X))$ $Orbit(A, \overline{b})$ (11) is satisfies to relations $S(\overline{b}) \to L(\overline{a}, \overline{c}, \overline{b}, \overline{b})$, $U(\overline{b}^{(m-1)}, \overline{b}) \to L(\overline{a}, \overline{c}, \overline{b}^{(m)}, \overline{b}),\ m = 1,2,...,m^*(\overline{b})$.

Remark 3. If the loop (10) is not completed (is branched) at some point $\overline{b}$, $m^*(\overline{b})$ it should be considered equal to infinity.: $m^*(\overline{b}) = +\infty$.

Example 7.

$S(x, y) = (0 \le x \le 1) \,\&\, (0 \le y \le 1)$,

$U(x, y, b_1, b_2) = \neg(|\, x + b_1 \,| \le \varepsilon) \,\&\, (|\, y + b_2 \,| \le \varepsilon)$,

$$A = \begin{bmatrix} 3/5 & 4/5 \\ -4/5 & 3/5 \end{bmatrix}.$$

$L = x + y \le 2b_1 + 2b_2 \,//\, \overline{a} = (1,1),\ \overline{c} = (2,2)$.

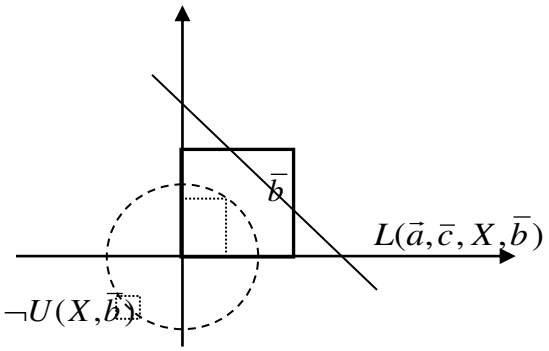


Fig. 2. Geometric illustration of the linear defined loop.

In this example, the linear operator A is an operator of rotation for angle $\alpha = arctg(4/3)$. A starting point $\bar{b}$ belongs to the unit square. The orbit of a linear operator A is a sequence, each point of which lies on the loop $x^2 + y^2 = b_1^2 + b_2^2$. The condition of repeating of the loop is a «point (x, y) that lies outside the square with side 2ε and center at $(-b_1,-b_2)$». Therefore, a loop is completed when the point gets inside this square, i.e. a point will make the rotation by angle $\pi + 2k\pi$ with accuracy equal to ε. Since the angle α is incommensurate with π, the orbit of the operator A is a dense set on the circle $x^2 + y^2 = b_1^2 + b_2^2$, therefore, the loop is complete. In this example, the basic algorithm is used to prove that $L = x + y \le 2b_1 + 2b_2$ is a conditional invariant of loop.

Let $f(x)$ be a minimal characteristic polynomial of the operator A, $\Lambda = \{\lambda_1,...,\lambda_n\}$ is a set of its roots (spectrum A). Suppose further that, $\lambda_1,...,\lambda_{2k}$ is a set of complex eigenvalues, and $\lambda_{2k+1},...,\lambda_n$ is a set of real eigennumbers and $\lambda_1 = \overline{\lambda}_2,...,\lambda_{2k-1} = \overline{\lambda}_{2k}$ than we obtain a representation of a linear operator in the so-called real Jordan form:

$$A' = \begin{bmatrix} B_1 & 0 & . & 0 & . & . & 0 \\ 0 & B_2 & . & 0 & . & . & 0 \\ . & . & . & . & . & . & . \\ 0 & ... & 0 & B_k & . & ... & 0 \\ 0 & . & . & 0 & \lambda_{2k+1} & ... & 0 \\ . & . & . & . & . & . & . \\ 0 & . & . & . & 0 & ... & \lambda_n \end{bmatrix}$$

Where $B_j = r_j \begin{bmatrix} \alpha_j & \beta_j \\ -\beta_j & \alpha_j \end{bmatrix}$.

Remark 4. After the transition to a basis of eigenvectors the coefficients of inequality will be changed. If $S(\Lambda)$ is a transition matrix, then the new values of the vectors $\bar{a}, \bar{b}$ calculated by the formulas $\bar{a}^{(S)} = S\bar{a}S^{-1}, \bar{b}^{(S)} = S\bar{b}S^{-1}$. But in order not to overload the text by new notations, we will use the old notations.

Note, that the matrix of the form $B \overset{df}{=} \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}$, where $\alpha^2 + \beta^2 = 1$ is a matrix of rotation of vector of two-dimensional space on the angle φ, that is defined by ratios $\cos(\varphi) = \alpha, \sin(\varphi) = \beta$. That is why

$$B_j = r_j \begin{bmatrix} \cos(\varphi_j) & \sin(\varphi_j) \\ -\sin(\varphi_j) & \cos(\varphi_j) \end{bmatrix}, \; r_j = \left|\lambda_j\right| = \sqrt{\alpha_j{}^2 + \beta_j{}^2}\,.$$

inequality (12), whose invariance is regarded by a loop (11) with a specific initial value $\bar{b}$, indicates that $\forall X \in Orbit(A, \bar{b})(\bar{a}, X) \le (\bar{c}, \bar{b})$. Algorithm of prove of the invariance of (12) will be formulated in the equivalent form:

$$\underset{X \in Orbit(A,\bar{b})}{Sup} (\bar{a}, X) \le (\bar{c}, \bar{b})\,.$$

Let us consider the linear form $a_1x_1 + a_2x_2 + \ldots + a_nx_n \overset{df}{=} (\bar{a}, X)$. The transformation $X := A * X$ converting this form in (a, AX), and m is a multiple iteration of loop, that is described by the transformation $X := A^m * X$ - in $(a, A^m X)$.

Let $X_1 = (x_1, x_2), \ldots, X_k = (x_{2k-1}, x_{2k})$, $\bar{a}_1 = (a_1, a_2), \ldots, \bar{a}_k = (a_{2k-1}, a_{2k})$. Then

$$(\bar{a}, X) = (\bar{a}_1, X_1) + \ldots + (\bar{a}_k, X_k) + a_{2k+1}x_{2k+1} + \ldots + a_nx_n \qquad (13)$$

Conversion $(\bar{a}, AX)$ of a linear form can be written as

$$(\bar{a}, AX) = (\bar{a}_1, B_1X_1) + \ldots + (\bar{a}_k, B_kX_k) + \lambda_{2k+1}a_{2k+1}x_{2k+1} + \ldots + \lambda_n a_n x_n \qquad (14)$$

And its m-th iteration can be written as

$$(\bar{a}, A^m X) = (\bar{a}_1, B_1^m X_1) + \ldots + (\bar{a}_k, B_k^m X_k) + \lambda_{2k+1}^m a_{2k+1}x_{2k+1} + \ldots + \lambda_n^m a_n x_n \qquad (15)$$

Passing in (14) to the representation in the form $B_j = r_j B_j$, we obtain:

$$(\bar{a}, A^m X) = r_1^m(\bar{a}_1, B_1^m X_1) + \ldots + r_k^m(\bar{a}_k, B_k^m X_k) + \lambda_{2k+1}^m a_{2k+1}x_{2k+1} + \ldots + \lambda_n^m a_n x_n$$

Consider the question of the set of values of the operator orbit $(\overline{a}_1, B_1^m X_1) + ... + (\overline{a}_k, B_k^m X_k)$ for the initial value $\overline{b}^{(0)} = (\overline{b}_1^{(0)},...,\overline{b}_k^{(0)})$, where $\overline{b}_j = (b_{2j-1}, b_{2j})$, $j = 1,...,k$. The interpreted pair X_j shall be as points on the two-dimensional plane, and the conversion of $\widehat{B}_j \overset{df}{=} \begin{bmatrix} \cos(\varphi_j) & \sin(\varphi_j) \\ -\sin(\varphi_j) & \cos(\varphi_j) \end{bmatrix}$ - as a rotations of points X_j on the angle φ_j.

The proof is formulated in lemmas 1-7 in [20].

Theorem 10. The problem of proving the invariance of inequality $L(\overline{a}, \overline{c}, X, \overline{b})$ for the loop (9) with diagonalizable linear operator A and with an initial point $\overline{b}$ is algorithmically solvable.

Theorem 11. The problem of proving the invariance of inequality $L(\overline{a}, \overline{c}, X, \overline{b})$ for the loop (9) (i.e., with the precondition $S(b)$) is algorithmically solvable.

Theorem 12. The problem of termination of the loop (9) is algorithmically solvable. Proof of theorems 10-12 is in [20].

4 Conclusion

This review represents main results of several works of one of the authors of the theory of program invariants. Subject of the research is an invariant of linear iteration loops. A new approach to the problems of static analysis of linear loops is represented: the problem of generating of polynomial invariance equations and the problem of proving the invariance of linear inequalities. This approach uses the representation of a linear operator in the loop body in the Jordan form and is based on the analysis of the spectrum of this operator.

The main results about invariant equality are the theorem 2 about multiplicative relations, a formula of invariant equations for the Jordan cell, a theorem 6 of the structure of a basis of the ideal of polynomial invariants, and, also, the algorithm of constructing of the basis of ideal of polynomial invariants for operators with irreducible over the field of rational numbers characteristic polynomial. Thus, for a given problem the problem of constructing of the basis of ideal of polynomial invariants for operators with a reducible characteristic polynomial remains open. From the practical view, the interest is in constructing the corresponding effective algorithms.

Unlike polynomial equations, the set of linear invariant inequalities does not have a finite basis. Therefore, a method of generating the basis is not applicable to this task. This paper represents the basic idea of the direct method of proof of the invariance of linear inequalities. There is a need to note, that the key role in the method is played by the set of maximal (from the modulus) eigenvalues of operator A. In this case, the case of maximal real eigenvalues and the maximal complex eigenvalues are significantly different. In the second case, the method uses the original method of

finding the maximum of the linear form in the orbit of a linear operator, and various algorithms of computation in the field of algebraic number.

There is a need to assume that this method can be used as a basis for a general algorithm of proving the invariance of a system of linear inequalities for linear-certain programs, similar to the method of proof of invariance of polynomial equations [5, 6], and to prove the invariance of polynomial inequalities for linear-certain programs.

References

1. Floyd, R.: Assigning Meanings to Programs. In: Proceedings of Symposium on Applied Mathematics, J.T. Schwartz (Ed.), American Mathematical Society, vol. 19, pp. 19--32, Providence, R.I. (1967)
2. Hoare, C.: An Axiomatic Basis for Computer Programming. Communications of the ACM 12(10), 576--580 (1969)
3. Letichevsky, A..: About One Approach to Program Analysis. Cybernetics 6, 1--8 (1979)
4. Godlevsky, A., Kapitonova, Y., Krivoy, S., Letichevsky, A.: Iterative Methods of Program Analysis. Cybernetics 2, 9--19 (1989)
5. Letichevsky, A., Lvov, M.: Discovery of Invariant Equalities in Programs over Data Fields. Applicable Algebra in Engineering, Communication and Computing 4, 21--29 (1993)
6. Müller-Olm, M., Seidl, H.: Precise Interprocedural Analysis Through Linear Algebra. In: Proc. of Symposium on Principles of Programming Languages, pp. 330--341, ACM, New York (2004)
7. Lvov M.: About One Algorithm of Program Polynomial Invariants Generation. Technical report, RISC Report Series (2007) (electronic).
8. Müller-Olm, M., Seidl, H.: Computing Polynomial Program Invariants. Inf. Process. Lett. 91(5), 233--244 (2004)
9. Sankaranarayanan, S., Sipma, H., Manna, Z.: Non-linear Loop Invariant Generation Using Gröbner Bases. In: Proc. of Symposium on Principles of Programming Languages, pp. 318--329, ACM, New York (2004)
10. Caplain, M.: Finding Invariant Assertions for Proving Programs. In: Proc. of the intern. Conf. on Reliable Software, pp. 165--171, ACM, New York (1975)
11. Rodríguez-Carbonell, E., Kapur, D.: Automatic Generation of Polynomial Loop Invariants: Algebraic Foundations. In: Proc. Of International Symposium on Symbolic and Algebraic Computation, pp. 266--273, ACM, New York (2004)
12. Rodríguez-Carbonell, E., Kapur, D.: Automatic Generation of Polynomial Invariants of Bounded Degree Using Abstract Interpretation. Sci. Comput. Program 64(1), 54--75 (2007)
13. Lvov, M.: A Method of Proving the Invariance of Linear Inequalities for Linear Loops. Cybernetics and Systems Analysis 4, 80--85 (2014)
14. Kovács, L. I., Jebelean, T.: An Algorithm for Automated Generation of Invariants for Loops with Conditionals. In: Proc. of Intern. Symposium on Symbolic and Numeric Algorithms for Scientific Computing. pp. 245--249, IEEE Computer Society, Timisoara (2005)
15. Kurosh, A.: Theory of Groups. 3-rd ed. Science, Moscow (1967)
16. Postnikov, M.: Galois Theory. Fizmatgiz, Moscow (1963)
17. Buchberger, B:. Gröbner Bases. An Algorithmic Method in the Theory of Polynomial Ideals. Computer algebra. Symbolic and algebraic computations. Mir, Moscow (1986)
18. Van Der Waerden: Algebra, B. the 2-th edition. GRFML, Moscow (1979)
19. Dieudonné, J. Carroll, Dj. Mumford, D.: Geometric Invariant Theory. Mir, Moscow (1974)

20. Lvov, M.: Analysis of Linear Defined Iterative Loops. Cybernetics and Systems Analysis 4 (2015) (In print)
21. Lvov, M.: Polynomial Invariants for Linear Loops. Cybernetics and Systems Analysis 4, 159--168 (2010)
22. Hodge, V., Pido, D.: Methods of Algebraic Geometry, Moscow (1954)
23. Lvov, M., Kreknin, V.: Nonlinear Invariants for Linear Loops and Eigenpolynomials of Linear Operators. Cybernetics and Systems Analysis 2, 126--139 (2012)
24. Kreknin, V., Lvov, M.: Eigenpolynomials of Linear Operators and Polynomial Invariants of Linear Loops of Program. Scientific Journal NEA Dragomanov 1(11), 150—169 (2010)
25. Lvov, M.: On the Structure of Polynomial Invariants of Linear Loops. (In print)
26. Cousot P., Halbwachs N.: Automatic Discovery of Linear Restraints among Variables of a Program. In: Conference Record of the Fifth Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, pp. 84--97, ACM Press, New York (1978)
27. Krivoy, S., Raksha, S.: Search of Invariant Linear Dependencies in Programs. Cybernetics 6, 23--28 (1984)
28. Godlewski, A., Kapitonova, Y, Krivoy, S., Letichevsky, A.: Iterative Methods of Programs Analysis. Equalities and Inequalities. Cybernetics 3, 1--10 (1990)
29. Lvov, M.: Invariant Inequalities in Programs Interpreted over an Ordered Field. Cybernetics 5, 22--27 (1986)
30. Lvov, M.: About Invariant Inequalities for States of the Program Schemes, that Interpreted Over the Vector Space. Cybernetics 2, 111--112 (1985)
31. Lvov, M.: A Method of Proving the Invariance of Linear Inequalities for Linear Loops. Cybernetics and Systems Analysis 4, 80--85 (2014)

Defining Finitely Supported Mathematics over Sets with Atoms

Andrei Alexandru and Gabriel Ciobanu

Romanian Academy, Institute of Computer Science, Iaşi
andrei.alexandru@iit.academiaromana-is.ro
gabriel@info.uaic.ro

Abstract. This paper presents some steps of defining a finitely supported mathematics by using sets with atoms. Such a mathematics generalizes the classical Zermelo-Fraenkel mathematics, and represents an appropriate framework to work with (infinite) structures in terms of finitely supported objects. We focus on the techniques of translating the Zermelo-Fraenkel results to this finitely supported mathematics over sets with atoms.

Keywords: Fraenkel-Mostowski set theory, invariant sets, finite support principle, Finitely Supported Mathematics.

Key-Terms: FormalMethod, MathematicalModel, Research.

1 Introduction

Since the experimental sciences are mainly interested in quantitative aspects, and since there exists no evidence for the presence of infinite structures, it becomes useful to develop a mathematics which deals with a more relaxed notion of (in)finiteness. We present our attempt of building the necessary concepts and structures for a finitely supported mathematics. What we call Finitely Supported Mathematics is a mathematics which is consistent with the axioms of the Fraenkel-Mostowski (FM) set theory. The FM axioms represents an "axiomatization" of the FM permutation model of the Zermelo-Fraenkel set theory with atoms; in this way, these axioms transform this model into an independent set theory. The axioms of the FM set theory are precisely the Zermelo-Fraenkel with atoms (ZFA) axioms over an infinite set of atoms [16], together with the special property of finite support which claims that for each element x in an arbitrary set we can find a finite set supporting x. Therefore in the FM universe only finitely supported objects are allowed. The original purpose of the FM set theory was to provide a mathematical model for variables in a certain syntax. Since they have no internal structure, atoms can be used to represent names. The finite support axiom is motivated by the fact that syntax can only involve finitely many names. The FM set theory provides a balance between rigorous formalism and informal reasoning. This is discussed in [23], where principles of

structural recursion and induction are explained in the FM framework. We can use this theory in order to manage infinite structures in a finitary manner, that is, in the FM framework we try to model the infinite using a more relaxed notion of finite, i.e, the notion of finite support.

Although a set of axioms for describing sets with atoms (or FM-sets) was introduced in [16], an earlier idea of using atoms in computer science belongs to Gandy [17]. Gandy proved that any machine satisfying four physical 'principles' is equivalent to some Turing machine. Gandy's four principles define a class of computing machines, namely the 'Gandy machines'. Gandy machines are represented by classes of 'states' and 'transition operations between states'. States are represented by hereditary finite sets built up from an infinite set U of atoms, and transformations are given by restricted operations from states to states. The class HF of all hereditary finite sets over U introduced in Definition 2.1 from [17] is described quite similar to the von-Neumann cumulative hierarchy of FM-sets, FM_A presented in [16]. The single difference between these approaches is that each HF_{n+1} is defined inductively involving '*finite* subsets of $U \cup HF_n$', whilst each $FM_{\alpha+1}(A)$ is defined inductively by using 'the disjoint union between A and the *finitely supported* subsets of $FM_\alpha(A)$'; HF is the union of all HF_n (with the mention that the empty set is not used in this construction), and the family of all FM-sets is the union of all FM_α from which we exclude the set A of atoms. The support of an element x in HF, obtained according to Definition 2.2(1) of [17], coincides with $supp(x)$ (with notations from Definition 2(4)) if we see x as an FM-set. Also, the effect of a permutation π on a structure x described in Definition 2.3 from [17] is defined analogue as the application of the S_A-action on FM_A to the element $(\pi, x) \in S_A \times FM_A$. Obviously, the Gandy's principles can also be presented in the FM framework because any finite set is well defined in FM; however, an open problem regards the consistency of Gandy's principles when 'finite' is replaced by 'finitely supported'.

The construction of the universe of all FM-sets [16] is inspired by the construction of the universe of all admissible sets over an arbitrary collection of atoms [6]. The hereditary finite sets used in [17] are particular examples of admissible sets. The FM-sets represent a generalization of hereditary finite sets because any FM-set is an hereditary finitely supported set.

In the literature there exist various approaches regarding the FM framework. We try to clarify the differences between these approaches.

- **The FM permutation model of the ZFA set theory.**
  This model was introduced by Fraenkel [14] and extended by Lindenbaum and Mostowski [21]. Its original aim was to establish the independence of the axiom of choice from the other axioms of the ZFA set theory. There also exist some other permutation models of ZFA presented in [20] which are defined by using countable infinite sets of atoms.
- **The FM axiomatic set theory.** This set theory was presented in [16]. It is inspired by the FM permutation model of the ZFA set theory. However, the FM set theory, the ZFA set theory and the Zermelo-Fraenkel (ZF) set theory are independent axiomatic set theories. All of these theories are described by

axioms, and all of them have models. For example, the Cumulative Hierarchy Fraenkel-Mostowski universe FM_A presented in [16] is a model of the FM set theory, while some models of the ZF set theory can be found in [19], and the permutation models of the ZFA set theory can be found in [20]. The sets defined using the FM axioms are called FM-sets. A ZFA set is an FM-set if and only if all its elements have hereditarily finite supports. Note that the infinite set of atoms in the FM set theory does not necessary be countable. The Fraenkel-Mostowski set theory is consistent whether the infinite set of atoms is countable or not. In [16] it is used a countable set of atoms in order to define a model of the Fraenkel-Mostowski set theory for new names in computer science, while in [7] there are described FM-sets over a set of atoms which do not represent a homogeneous structure. Also, in [12] the authors use non-countable sets of atoms (like the set of real numbers) in order to study the minimization of deterministic timed automata.

– **Nominal sets.** These sets can be defined both in the ZF framework [24] and in the FM framework [16]. In ZF, a fixed infinite set A is considered as a set of names. A nominal set is defined as a usual ZF set endowed with a particular group action of the group of permutations over A that satisfies a certain finiteness property. Such a finiteness property allows us to say that nominal sets are well defined according to the axioms of the FM set theory whenever the set of names is the set of atoms in the FM set theory. There exists also an alternative definition for nominal sets in the FM framework. They can be defined as sets constructed according to the FM axioms with the additional property of being empty supported (invariant under all permutations). These two ways of defining nominal sets finally lead to similar properties. According to the previous remark we use the terminology "invariant" for "nominal" in order to establish a connection between approaches in the FM framework and in the ZF framework. Moreover, we can say that any set defined according to the FM axioms (any FM-set) can be seen as a subset of the nominal (invariant) set FM_A. However, an FM-set is itself a nominal set only if it has an empty support. The theory of nominal sets makes sense even if the set of atoms is infinite but not countable. Informally, since the ZFA set theory collapses into the ZF set theory when the set of atoms is empty, we can say that the nominal sets represent a natural extension of the usual sets. In computer science, nominal sets offer an elegant formalism for describing λ-terms modulo α-conversion [16]. They can also be used in algebra [5, 2], in proof theory [27], in domain theory [26], in topology [22], semantics of process algebras [4, 15] and programming [25]. A survey on the applications of nominal sets in computer science emphasizing our contributions can be found in [3].
– **Generalized nominal sets.** The theory of nominal sets over a fixed set A of atoms is generalized in [10] to a new theory of nominal sets over arbitrary (unfixed) sets of data values. This provides the generalized nominal sets. The notion of 'S_A-set' (Definition 2) is replaced by the notion of 'set endowed with an action of a subgroup of the symmetric group of $\mathbb{D}$' for an arbitrary set of data values $\mathbb{D}$, and the notion of 'finite set' is replaced by the notion of 'set

with a finite number of orbits according to the previous group action (orbit-finite set)'. This approach is useful for studying automata on data words [10], languages over infinite alphabets [8], or Turing machines that operate over infinite alphabets [11]. Computations in these generalized nominal sets are presented in [9, 13].

As their names say, the nominal sets are used to manage notions like renaming, binding or fresh name. However, this theory could be studied deeper from an algebraically viewpoint, and it could be used in order to characterize some infinite structures in terms of finitely supported objects.

Finitely Supported Mathematics (FSM) is introduced to prove that many finiteness ZF properties still remain valid if we replace the term 'finite' with 'infinite, but with finite support'. Such results have already been presented in [5] where we proved that a class of multisets over infinite alphabets (interpreted in the nominal framework) has similar properties to the classical multisets over finite alphabets. FSM is the mathematics developed in the world of finitely supported objects where the set of atoms has to be infinite (countable or not countable). Informally, FSM extends the framework of the ZF set theory without choice principles; ZF set theory is actually the Empty Supported Mathematics. In FSM, we use either 'invariant sets' or 'finitely supported sets' instead of 'sets'. As an intuitive rule, we are not allowed to use in the proofs of the results of FSM any construction that does not preserve the property of finite support. That means we cannot obtain a property in FSM only by using a ZF result without an appropriate proof using only the finite support condition. Since the invariant sets can also be defined in the ZFA framework similarly as in the ZF framework (see the first paragraph in Section 2), the definition of the finitely supported mathematics also makes sense over the ZFA axioms.

To summarize, FSM represents the ZF theory rephrased in terms of finitely supported objects; this means that FSM presents the theory of invariant sets, including invariant algebraic structures. FSM is not at all the theory of nominal sets from [24] presented in a different manner; actually the theory of nominal sets [24] could be considered as a tool for defining FSM. The main aim of FSM is to characterize the infinite algebraic structures by using their finite supports.

2 Sets with Atoms

Let A be a fixed infinite (countable or non-countable) ZF-set. The following results make also sense if A is considered to be the set of atoms in the ZFA framework (characterized by the axiom "$y \in x \Rightarrow x \notin A$") and if 'ZF' is replaced by 'ZFA' in their statements. Thus, we mention that the theory of invariant sets makes sense both in ZF and in ZFA. Several results of this section are similar to those in [24], but without assuming the set of atoms to be countable.

Definition 1. *A* transposition *is a function* $(a\,b) : A \to A$ *defined by* $(a\,b)(a) = b$, $(a\,b)(b) = a$, *and* $(a\,b)(n) = n$ *for* $n \neq a, b$. *A* permutation *of A is generated by composing finitely many transpositions.*

Definition 2. *Let S_A be the set of all permutations of A.*

1. *Let X be a ZF set. An* S_A-action *on X is a function $\cdot : S_A \times X \to X$ having the properties that $Id \cdot x = x$ and $\pi \cdot (\pi^{'} \cdot x) = (\pi \circ \pi^{'}) \cdot x$ for all $\pi, \pi' \in S_A$ and $x \in X$. An* S_A-set *is a pair $(X, \cdot)$ where X is a ZF set, and $\cdot : S_A \times X \to X$ is an S_A-action on X.*
2. *Let $(X, \cdot)$ be an S_A-set. We say that $S \subset A$* supports *x whenever for each $\pi \in Fix(S)$ we have $\pi \cdot x = x$, where $Fix(S) = \{\pi \mid \pi(a) = a, \forall a \in S\}$.*
3. *Let $(X, \cdot)$ be an S_A-set. We say that X is an* invariant set *if for each $x \in X$ there exists a finite set $S_x \subset A$ which supports x. Invariant sets are also called* nominal sets *if we work in the ZF framework [24], or* equivariant sets *if they are defined as elements in the cumulative hierarchy FM_A [16].*
4. *Let X be an S_A-set and let $x \in X$. If there exists a finite set supporting x, then there exists a least finite set supporting x [16] which is called* the support of *x and is denoted by $supp(x)$. An element supported by the empty set is called equivariant.*

Proposition 1. *Let $(X, \cdot)$ be an S_A-set and $\pi \in S_A$. If $x \in X$ is finitely supported, then $\pi \cdot x$ is finitely supported, and $supp(\pi \cdot x) = \pi(supp(x))$.*

Example 1.

1. The set A of atoms is an S_A-set with the S_A-action $\cdot : S_A \times A \to A$ defined by $\pi \cdot a := \pi(a), \forall \pi \in S_A, a \in A$. Moreover, $supp(B) = B, \forall B \subset A$, B finite.
2. Any ordinary ZF set X (like $\mathbb{N}$ or $\mathbb{Z}$) is an S_A-set with the trivial S_A-action $\cdot : S_A \times X \to X$ defined by $\pi \cdot x := x$ for all $\pi \in S_A$ and $x \in X$.
3. If $(X, \cdot)$ is an S_A-set, then $\wp(X) = \{Y \mid Y \subseteq X\}$ is also an S_A-set with the S_A-action $\star : S_A \times \wp(X) \to \wp(X)$ defined by $\pi \star Y := \{\pi \cdot y \mid y \in Y\}$ for all $\pi \in S_A$, and all subsets Y of X. For each invariant set $(X, \cdot)$ we denote by $\wp_{fs}(X)$ the set formed from those subsets of X which are finitely supported according to the action $\star$. $(\wp_{fs}(X), \star|_{\wp_{fs}(X)})$ is an invariant set, where $\star|_{\wp_{fs}(X)}$ represents the action $\star$ restricted to $\wp_{fs}(X)$.
4. Let $(X, \cdot)$ and $(Y, \diamond)$ be S_A-sets. The Cartesian product $X \times Y$ is also an S_A-set with the S_A-action $\star : S_A \times (X \times Y) \to (X \times Y)$ defined by $\pi \star (x, y) = (\pi \cdot x, \pi \diamond y)$ for all $\pi \in S_A$ and all $x \in X$, $y \in Y$. If $(X, \cdot)$ and $(Y, \diamond)$ are invariant sets, then $(X \times Y, \star)$ is also an invariant set.
5. The FM cumulative hierarchy FM_A described in [16] is an invariant set with S_A-action $\cdot : S_A \times FM_A \to FM_A$ defined inductively by $\pi \cdot a := \pi(a)$ for all atoms $a \in A$ and $\pi \cdot x := \{\pi \cdot y \mid y \in x\}$ for all $x \in FM_A \setminus A$. An FM-set is a finitely supported element in FM_A; additionally an FM-set has the recursive property that all its elements are also FM-sets. An FM-set which is empty supported as an element in FM_A is an invariant set.

Definition 3. *Let $(X, \cdot)$ be an invariant set. A subset Z of X is called* finitely supported *if and only if $Z \in \wp_{fs}(X)$ with the notations of Example 1 (3).*

Definition 4. *Let X and Y be invariant sets, and let Z be a finitely supported subset of X. A function $f : Z \to Y$ is* finitely supported *if $f \in \wp_{fs}(X \times Y)$.*

Proposition 2. *[5] Let $(X, \cdot)$ and $(Y, \diamond)$ be invariant sets, and let Z be a finitely supported subset of X. The function $f : Z \to Y$ is finitely supported in the sense of Definition 4 if and only if there exists a finite set S of atoms such that for all $x \in Z$ and all $\pi \in Fix(S)$ we have $\pi \cdot x \in Z$ and $f(\pi \cdot x) = \pi \diamond f(x)$.*

3 Reformulating the Classical ZF Results in FSM

The main idea of translating a classical ZF result (depending on sets and relations) into FSM is to analyze if there exists a valid result obtained by replacing "set" with "invariant/finitely supported set" and "relation" with "invariant/finitely supported relation" in the ZF result. If this is possible, then things go smoothly; however, this is not always so simple.

Every ZF set is a particular invariant set equipped with a trivial permutation action (Example 1(2)). Therefore, the general properties of invariant sets lead to valid properties of ZF sets. The converse is not always valid, namely not every ZF result can be directly rephrased in the world of invariant sets, terms of finitely supported objects according to arbitrary permutation actions. This is because, given an invariant set X, there could exist some subsets of X (and also some relations or functions involving subsets of X) which fail to be finitely supported. A classical example (presented also in Subsection 2.2.3.6 of [26]) is represented by the powerset of the invariant set A. A subset of A which is in the same time infinite and coinfinite could be defined in some models of ZF (or of ZFA is we consider A to be the set of atoms in ZFA), but it can not be defined in FSM because it is not finitely supported. Therefore the remark that everything that can be done in ZF can also be done in FSM is not valid. That means there may exist some valid results depending on several ZF structures which fail to be valid in FSM if we simply replace "ZF structure" with "FSM structure" in their statement.

We present few examples regarding these aspects. There exist some valid ZF results that cannot be translated into FSM. According to Remark 1, the following examples are particularly interesting because they do not overlap neither on some known properties of permutative models of ZFA, nor on some properties of nominal sets [24].

Example 2.

- There exist models of ZF without choice that satisfy the ordering principle "Every set can be totally ordered". More details about such models are in [19], where there are mentioned Howard-Rubin's first model N38 and Cohen's first model M1. Therefore the ordering principle is independent from the axioms of the ZF set theory.
- In FSM the following result fails "For every invariant set X there exists a finitely supported total order relation on X". Therefore the ordering principle is inconsistent with the axioms of the FM set theory. Indeed, suppose that there exists a finitely supported total order $<$ on the invariant set A. Let $a, b, c \notin supp(<)$ with $a < b$. Since $(a\, c) \in Fix(supp(<))$ we have $(a\, c)(a) <$

$(a\,c)(b)$, so $c < b$. However, we also have $(a\,b), (b\,c) \in Fix(supp(<))$, and so $((a\,b) \circ (b\,c))(a) < ((a\,b) \circ (b\,c))(b)$, that is, $b < c$. We get a contradiction, and so the translation of the ordering principle in FSM realized by replacing "structure" with "finitely supported structure" leads to a false statement.

Example 3.

- There exist models of ZF without choice that satisfy the partial countable choice principle: "Given any countable family (sequence) of non-empty sets $\mathcal{F} = (X_n)_n$, there exists an infinite subset M of $\mathbb{N}$ such that it is possible to select a single element from each member of the family $(X_m)_{m\in M}$, i.e. there exist a choice function on $(X_m)_{m\in M}$". More details about such models are in [19], where there are mentioned Pincus-Solovay's First Model M27, Shelah's Second Model M38 and Howard-Rubin's first model N38. Therefore the partial countable choice principle is independent from the axioms of the ZF set theory.
- In FSM the following result fails: "Given any invariant set X, and any countable family $\mathcal{F} = (X_n)_n$ of subsets of X such that the mapping $n \mapsto X_n$ is finitely supported, there exists an infinite subset M of $\mathbb{N}$ with the property that there is a finitely supported choice function on $(X_m)_{m\in M}$". Therefore the partial countable choice principle is inconsistent with the axioms of the FM set theory. Indeed, for the invariant set A we consider the countable family $(X_n)_n$ where X_n is the set of all injective n-tuples from A. Since A is infinite, it follows that each X_n is non-empty. In the FM framework, each X_n is equivariant because A is an invariant set and each permutation is a bijective function. Therefore the family $(X_n)_n$ is equivariant, and the mapping $n \mapsto X_n$ is also equivariant. Suppose that there exists an infinite subset M of $\mathbb{N}$ and a finitely supported choice function f on $(X_m)_{m\in M}$. Let $f(X_m) = y_m$ with each $y_m \in X_m$. Let $\pi \in Fix(supp(f))$. According to Proposition 2, and because each element X_m is equivariant according to its definition, we obtain that $\pi \cdot y_m = \pi \cdot f(X_m) = f(\pi \cdot X_m) = f(X_m) = y_m$. Therefore, each element y_m is supported by $supp(f)$, and so $supp(y_m) \subseteq supp(f)$ for all $m \in M$. Since y_m is a finite tuple of atoms which has exactly m elements for each $m \in M$, we have that $supp(y_m) = y_m$, $\forall m \in \mathbb{N}$ (see Example 1(1)). Thus $y_m \subseteq supp(f)$ for all $m \in M$. However, because M is infinite, we contradict the finiteness of $supp(f)$. Therefore the translation of the partial countable choice principle in FSM realized by replacing "structure" with "finitely supported structure" leads to a false statement.

Remark 1. Examples 2 and 3 show us that there exist some choice principles which are independent from the axioms of the ZF set theory, but inconsistent in FSM. Since FSM is consistent even if the set of atoms is not countable, such results do not overlap on some related properties in the basic or in the second Fraenkel modes of the ZFA set theory (which are defined using countable sets of atoms) [20]. Also, the previous results do not follow immediately from [24] because the nominal sets are defined over countable sets of atoms, while we define invariant sets over possible non-countable sets of atoms; in [24] where

the set of atoms is countable, Example 3 would be trivial. Moreover, we claim that all the choice principles from [18] rephrased in terms of invariant sets are inconsistent in FSM. Note that it is not easy to prove such a result in FSM, even if various relationship results between several forms of choice hold in the ZF framework. This is because nobody guarantees that ZF results remain valid in FSM. Therefore, all the possible relationship results between various choice principles in FSM have to be independently proved in terms of finitely supported object. Details regarding the consistency of various choice principles in the world of invariant sets defined over possibly non-countable sets of atoms are presented in another paper.

Other results which fail in FSM are given by the Stone duality [22], by the determinization of finite automata and by the equivalence of two-way and one-way finite automata [10]. There also exist some valid ZF results that can be translated into FSM only in a weaker form.

Example 4. We define an invariant complete lattice as an invariant set $(L, \cdot)$ together with an equivariant order relation $\sqsubseteq$ on L satisfying the property that every finitely supported subset $X \subseteq L$ has a least upper bound with respect to the order relation $\sqsubseteq$.

- Let L be a ZF complete lattice and $f : L \to L$ a ZF monotone function. Then there exists a greatest $e \in L$ such that $f(e) = e$ and a least $e \in L$ such that $f(e) = e$ (weak form of Tarski theorem).
- Let $(L, \sqsubseteq, \cdot)$ be an invariant complete lattice and $f : L \to L$ a finitely supported monotone function. Then there exists a greatest $e \in L$ such that $f(e) = e$, and a least $e \in L$ such that $f(e) = e$ (the proof is similar to Theorem 3.2 in [1]).

These results show that the weak form of the Tarski theorem can be naturally translated into FSM. However, as it is presented below, the strong form of the Tarski theorem cannot be naturally translated into FSM; it holds in FSM only for a particular class of finitely supported monotone functions, i.e, the equivariant monotone functions.

- Let L be a ZF complete lattice and $f : L \to L$ a ZF monotone function over L. Let P be the set of fixed points of f. Then P is a complete lattice (strong form of Tarski theorem).
- Let $(L, \sqsubseteq, \cdot)$ be an invariant complete lattice and $f : L \to L$ an equivariant monotone function over L. Let P be the set of fixed points of f. Then $(P, \sqsubseteq, \cdot)$ is an invariant complete lattice.
  The result does not hold if f is finitely supported, but not equivariant (the proof is similar to Theorem 3.3 in [1]).

4 Limits of the Equivariance / Finite Support Principle

In order to translate a general ZF result into FSM, one must prove that several structures are finitely supported. There exist two general methods of proving

that a certain structure is finitely supported. The first method is a constructive one: by using some intuitive arguments, we anticipate a possible candidate for the support and prove that this candidate is indeed a support. The second method is based on a general finite support principle which is defined using the higher-order logic. However the use of this second method has some limits, as we present in the paragraphs below.

According to [23], we have the following equivariance/finite support principle which works over invariant sets.

Theorem 1.

- *Any function or relation that is defined from equivariant functions and relations using classical higher-order logic is itself equivariant.*
- *Any function or relation that is defined from finitely supported functions and relations using classical higher-order logic is itself finitely supported.*

In applying this equivariance/finite support principle, one must take into account all the parameters upon which a particular construction depends. We think that the formal involvement of the equivariance/finite support principle, i.e. the precise verification if the conditions for applying the equivariance/finite support principle are properly satisfied is sometimes at least as difficult as a constructive proof. Moreover, in many cases we need to construct effectively the support, and it is not enough to prove only that a certain structure is finitely supported.

Example 5. An invariant monoid $(M, \cdot, \diamond)$ is an invariant set $(M, \diamond)$ endowed with an equivariant internal monoid law $\cdot : M \times M \to M$. If $(\Sigma, \diamond)$ is an invariant set, then the free monoid Σ^* on Σ is an invariant monoid [5].

1. For each monoid M and each function $f : \Sigma \to M$, there exists a unique homomorphism of monoids $g : \Sigma^* \to M$ with $g \circ i = f$, where $i : \Sigma \to \Sigma^*$ is the standard inclusion of Σ into Σ^* which maps each element $a \in \Sigma$ into the word a (ZF universality theorem for monoids).
2. i) Let $(\Sigma, \diamond_\Sigma)$ be an invariant set. Let $i : \Sigma \to \Sigma^*$ be the standard inclusion of Σ into Σ^* which maps each element $a \in \Sigma$ into word a. If $(M, \cdot, \diamond_M)$ is an arbitrary invariant monoid and $\varphi : \Sigma \to M$ is an arbitrary finitely supported function, then there exists a unique finitely supported homomorphism of monoids $\psi : \Sigma^* \to M$ with $\psi \circ i = \varphi$.
   This result can be proved directly by involving the equivariance/finite support principle.
   ii) Let $(\Sigma, \diamond_\Sigma)$ be an invariant set. Let $i : \Sigma \to \Sigma^*$ be the standard inclusion of Σ into Σ^* which maps each element $a \in \Sigma$ into the word a. If $(M, \cdot, \diamond_M)$ is an arbitrary invariant monoid and $\varphi : \Sigma \to M$ is an arbitrary finitely supported function, then there exists a unique finitely supported homomorphism of monoids $\psi : \Sigma^* \to M$ with $\psi \circ i = \varphi$. Moreover, if a finite set S supports φ, then the same set S supports ψ.
   The last sentence of this theorem cannot be proved by involving the equivariance/finite support principle.

Proof. If $(M,\cdot,\diamond_M)$ is an invariant monoid, then $(M,\cdot)$ is a monoid. From the general ZF theory of monoids, we can define a unique homomorphism of monoids $\psi:\Sigma^*\to M$ with $\psi\circ i=\varphi$.

In [5] we proved that the free monoid Σ^* on Σ is an invariant monoid whenever $(\Sigma,\diamond)$ is an invariant set. The S_A-action $\widetilde{\star}:S_A\times\Sigma^*\to\Sigma^*$ is defined by $\pi\widetilde{\star}x_1x_2\ldots x_l=(\pi\diamond x_1)\ldots(\pi\diamond x_l)$ for all $\pi\in S_A$ and $x_1x_2\ldots x_l\in\Sigma^*\setminus\{1\}$, and $\pi\widetilde{\star}1=1$ for all $\pi\in S_A$.

In order to prove that ψ is finitely supported it is sufficient to apply Theorem 1 because ψ is defined from the finitely supported functions φ and i using the higher-order logic. However, Theorem 1 is not sufficient to prove that if a finite set S supports φ, then the same set S supports ψ. In order to prove the previous statement we proceed as follows.

Let us consider $S=supp(\varphi)$. Thus, by Proposition 2 we have $\varphi(\pi\diamond_\Sigma x)=\pi\diamond_M\varphi(x)$ for all $x\in\Sigma$ and $\pi\in Fix(S)$. We have to prove that S supports ψ. Let $\pi\in Fix(S)$. According to Proposition 2 it is sufficient to prove that $\psi(\pi\widetilde{\star}x_1x_2\ldots x_n)=\pi\diamond_M\psi(x_1x_2\ldots x_n)$ for each $x_1x_2\ldots x_n\in\Sigma^*$. However, ψ is a monoid homomorphism between Σ^* and M, and $\psi\circ i=\varphi$. This means $\psi(x_1x_2\ldots x_n)=\varphi(x_1)\cdot\varphi(x_2)\cdot\ldots\cdot\varphi(x_n)$. Since $(M,\cdot,\diamond_M)$ is an invariant monoid we have $\pi\diamond_M\psi(x_1x_2\ldots x_n)=\pi\diamond_M(\varphi(x_1)\cdot\varphi(x_2)\cdot\ldots\cdot\varphi(x_n))=(\pi\diamond_M\varphi(x_1))\cdot(\pi\diamond_M\varphi(x_2))\cdot\ldots\cdot(\pi\diamond_M\varphi(x_n))=\varphi(\pi\diamond_\Sigma x_1)\cdot\varphi(\pi\diamond_\Sigma x_2)\cdot\ldots\cdot\varphi(\pi\diamond_\Sigma x_n)$. However, $\pi\widetilde{\star}x_1x_2\ldots x_n=(\pi\diamond_\Sigma x_1)\ldots(\pi\diamond_\Sigma x_n)$ and $\psi(\pi\widetilde{\star}x_1x_2\ldots x_n)=\psi((\pi\diamond_\Sigma x_1)\ldots(\pi\diamond_\Sigma x_l))=\varphi(\pi\diamond_\Sigma x_1)\cdot\varphi(\pi\diamond_\Sigma x_2)\cdot\ldots\cdot\varphi(\pi\diamond_\Sigma x_n)$. Hence $\psi(\pi\widetilde{\star}x_1x_2\ldots x_n)=\pi\diamond_M\psi(x_1x_2\ldots x_n)$ for each $\pi\in Fix(S)$, which means S supports ψ.

Example 5(2) shows us that by using the equivariance/finite support principle we can obtain a universality property for invariant monoids which is similar to the one described in Example 5(1). However, in order to prove that $supp(\psi)\subseteq supp(\varphi)$ in the second item of Example 5(2), we need to present a constructive method of defining a set supporting ψ (see also Theorem 6 from [5]). Other related examples regarding the equivariance/finite support principle are Theorems 4, 9 and 11 from [5], or Theorem 3.7 from [2]. In these theorems we are able to prove a precise characterization for the support of some structures which could not be obtained by a direct application of the equivariance/finite support principle in the form from Theorem 1. In these results we do not prove only that some structures are finitely supported, but we also found a relationship between the supports of the related structures.

In some cases we can prove stronger properties without involving the equivariance/finite support principle. For example, each function f_x in the proof of Theorem 7 of [5] has a non-empty finite support. Using the equivariance/finite support principle one can say that the function T from that theorem has also a finite support. We were able to prove something stronger using a constructive method: the function T is equivariant.

A *constructive method* of defining the support is also necessary in order to assure that some structures are uniformly finitely supported (i.e. supported by the same finite set of atoms). Some related examples regarding the uniform support are presented in [2] (Section 5), where we should assume that some

structures are uniformly supported in order to obtain some embedding properties for invariant (nominal) groups. Also, note that a chain is finitely supported if and only if all its elements are finitely supported and have the same support, i.e., all its elements are uniformly finitely supported. Therefore, in order to prove that a chain is finitely supported, we must present a constructive method of defining the support of its elements. More exactly, we cannot use the equivariance/finite support principle which would not assure the uniformity of the support of its elements. Suggestive examples regarding finitely supported chains are presented in Chapter 4 of [25].

We conclude that the equivariance/finite support principle is not useful when we want to obtain a relationship between the supports of several constructions (and we do not want only to prove that these constructions are finitely supported). This is because, in its actual form, the second part of Theorem 1 allows to prove that a certain structure is finitely supported, but it do not provide any information about the structure of the support. However, the first part of Theorem 1 helps when we want to prove the equivariance of some constructions. Note that we do not claim that the finite support principle is not useful. Obviously, it can be used to give simpler proofs for the fact that functions and relations defined from finitely supported functions and relations via classical higher-order formulas are finitely supported. However, a concrete calculation for the supports of some structures is able to provide more informations about the related supports; we justify this viewpoint in Example 5. Also, such a method is useful in order to find the uniform supports.

Note that, often in practice, it is not sufficient to prove only that a certain structure is finitely supported without giving any information about the structure of support. A more precise characterization of the support is useful. For example, let us consider an α-equivalence class $[t]$ of a λ-term t. The support of $[t]$ is represented by the set of free names of t [16]; the support of $[t]$ is finite because any λ-term has a finite numbers of free names. However, the precise description of the free names of t is an aspect that matters. Therefore, we suggest to use a constructive method of defining the support of a certain structure instead of the finite support part (the second part) of Theorem 1, because in this way we can obtain more informations about the support.

5 Conclusion

Our goal is to develop a mathematics for experimental science which deals with a more relaxed notion of finiteness. We call it the 'Finitely Supported Mathematics'. Informally, in Finitely Supported Mathematics we can model infinite structures after a finite number of observations. More precisely, we intend to restate some parts of algebra by replacing '(infinite) sets' with 'invariant sets'. This allows to model some infinite structures by using their finite supports. In order to sustain our viewpoint, we involve the axiomatic theory of FM-sets presented in [16]. Rather than using a non-standard set theory, we could alternatively work with invariant sets, which are defined within ZF as usual sets endowed with some

group actions satisfying a finite support requirement. The properties of invariant sets are similar to those presented in [24], with the mention that we assume invariant sets to be defined over possible non-countable sets of atoms. Our paper presents the basic steps requested in order to provide an extension of the theory of invariant sets to a theory of invariant algebraic structures. Although the initial purpose of defining invariant sets was to formulate a semantics for syntax with variable binding, we consider that such sets can also be used from an algebraic perspective in order to characterize infinite structures modulo finite supports, and thus in order to provide more informations about infinite objects.

The category of invariant sets has a very rich structure, and so the definitions of many structures given in the usual category of sets can be reformulated within the invariant sets framework. A natural question is which classical theorems about these structures hold internally in the world of invariant sets. Until now (or, more precisely, until we would be able to solve the open problem presented below), there does not exist a standard algorithm to translate any classical ZF result into FSM. This is because there may exist some subsets of an invariant set which fail to be finitely supported, and thus there may exist some ZF results that fail in the universe of invariant sets. Related examples regarding the previous statement are presented in Section 3. Therefore, reformulating the ZF theorems into FSM should be done for each case separately. For example, the theory of monoids is studied in FSM in [5], the theory of groups is rephrased in FSM in [2], and the theory of posets and domains is reformulated within invariant sets framework in [24, 25]. In order to prove that a structure is finitely supported, one could use either the finite support principle of [24] (e.g. Theorem 1), or a more "constructive" method. To employ such a "constructive method" means that we anticipate a possible candidate for a support, and then prove that this candidate is indeed a support. The benefit of this method is that we are able to obtain more informations about the related support than by using the finite support principle. Related examples can be found in Section 4.

An Open Problem: The main task in order to define a finitely supported mathematics is to prove that certain subsets of an invariant set are finitely supported. We already know that given an invariant set X, there could exist some subsets of X which fail to be finitely supported. Some related examples are presented in [24] and [26]. However, all these examples are described by using choice principles or consequences of choice principles (like the assertion that the set A can be non-amorphous in ZFA) in order to construct some structures which later fail to be finitely supported. We conjecture that all the choice principles presented in [18] are inconsistent in FSM. We did not find yet any example of a non-finitely supported subset of an invariant set defined without using a choice principle from [18] or a consequence of a form of choice (like the construction of an infinite and coinfinite subset of an infinite set). Therefore, the question regarding the validity of the following assertions naturally appears.

– If we consider the ZF set theory (or the ZFA set theory) without any choice principle, then every subset of an invariant set is finitely supported?

– For what kind of atoms the previous question has an affirmative answer?

If we get an affirmative answer (even for a particular set of atoms), then the mathematics developed in the ZF (or ZFA) set theory without any choice principle would be somehow equivalent to FSM, namely we could model any infinite structure by using its finite support.

Acknowledgements. The work was supported by a grant of the Romanian National Authority for Scientific Research, CNCS-UEFISCDI, project number PN-II-ID-PCE-2011-3-0919.

References

1. Alexandru, A., Ciobanu, G.: Nominal event structures. Romanian Journal of Information, Science and Technology. 15, 79–90 (2012)
2. Alexandru, A., Ciobanu, G.: Nominal groups and their homomorphism theorems. Fundamenta Informaticae. 131(3-4), 279–298 (2014)
3. Alexandru, A., Ciobanu, G.: On the development of the Fraenkel-Mostowski set theory. Bulletin Inst. Politehnic Iasi. LX, 77–91 (2014)
4. Alexandru, A., Ciobanu, G.: A nominal approach for fusion calculus. Romanian Journal of Information Science and Technology. 17, (2014)
5. Alexandru, A., Ciobanu, G.: Mathematics of multisets in the Fraenkel-Mostowski framework. Bulletin Mathematique de la Societe des Sciences Mathematiques de Roumanie. 58/106 (1), 3–18 (2015)
6. Barwise, J.: Admissible Sets and Structures: An Approach to Definability Theory, Perspectives in Mathematical Logic. Vol.7, Springer (1975)
7. Bojanczyk, M.: Fraenkel-Mostowski sets with non-homogeneous atoms. Lecture Notes in Computer Science. 7550, 1–5, Springer (2012)
8. Bojanczyk, M.: Nominal monoids. Theory of Computing Systems. 53, 194–222 (2013)
9. Bojanczyk, M., Braud L., Klin, B., Lasota, S.: Towards nominal computation. In: 39th ACM POPL, pp. 401–412 (2012)
10. Bojanczyk, M., Klin, B., Lasota, S.: Automata with group actions. In: 26th Symposium on Logic in Computer Science, pp. 355–364. IEEE Press (2011)
11. Bojanczyk, M., Klin, B., Lasota, S., Torunczyk, S.: Turing machines with atoms. In: 28th Symposium on Logic in Computer Science, pp.183–192. IEEE Press (2013)
12. Bojanczyk, M., Lasota, S.: A Machine-independent characterization of timed languages. In: 39th ICALP, 92–103 (2012)
13. Bojanczyk, M., Torunczyk, S.: Imperative programming in sets with atoms. In: FSTTCS. LIPIcs vol.18, pp. 4–15 (2012)
14. Fraenkel, A.: Zu den grundlagen der Cantor-Zermeloschen mengenlehre. Mathematische Annalen. 86, 230–237 (1922)
15. Gabbay, M.J.: The pi-calculus in FM. Thirty Five Years of Automating Mathematics, Kluwer Applied Logic. 28, pp. 247–269 (2003)
16. Gabbay, M.J., Pitts, A.M.: A new approach to abstract syntax with variable binding. Formal Aspects of Computing. 13, 341–363 (2001)
17. Gandy, R.: Church's thesis and principles for mechanisms, In: Barwise, J., Keisler, H.J., Kunen, K.(eds). The Kleene Symposium, pp. 123–148, North-Holland (1980)
18. Herrlich, H. Axiom of Choice. Lecture Notes in Mathematics. Springer (2006)

19. Howard, P., Rubin, J.E.: Consequences of the Axiom of Choice. Mathematical Surveys and Monographs vol.59. American Mathematical Society (1998)
20. Jech, T. J.: The Axiom of Choice. Studies in Logic and the Foundations of Mathematics. North-Holland (1973)
21. Lindenbaum, A., Mostowski, A.: Uber die unabhangigkeit des auswahlsaxioms und einiger seiner folgerungen. Comptes Rendus des Seances de la Societe des Sciences et des Lettres de Varsovie. 31, 27–32 (1938)
22. Petrisan, D.: Investigations into Algebra and Topology over Nominal Sets. PhD Thesis, University of Leicester (2011)
23. Pitts. A.M.: Alpha-structural recursion and induction. Journal of the ACM. 53, 459–506 (2006)
24. Pitts, A.M.: Nominal Sets Names and Symmetry in Computer Science. Cambridge University Press (2013)
25. Shinwell, M.R.: The Fresh Approach: Functional Programming with Names and Binders. PhD Thesis, University of Cambridge (2005)
26. Turner, D.: Nominal Domain Theory for Concurrency. Technical Report no.751, University of Cambridge (2009)
27. Urban, C.: Nominal techniques in Isabelle/HOL. Journal of Automated Reasoning. 40, 327–356 (2008)

On a Strong Notion of Viability for Switched Systems

Ievgen Ivanov

Taras Shevchenko National University of Kyiv, Ukraine
ivanov.eugen@gmail.com

Abstract. We propose a strong notion of viability for a set of states of a nonlinear switched system. This notion is defined with respect to a fixed region of the state space and can be interpreted as a condition under with a system can be forced to stay in a given safe set by applying a specific control strategy only when its state is outside the fixed region. When the state of the system is inside the fixed region, the control can be kept constant without the risk of driving the system into unsafe set (the complement of the safe set).
We investigate and give a convenient sufficient condition for strong viability of the complement of the origin for a nonlinear switched system with respect to a fixed region.

Keywords. dynamical system, switched system, viability, global-in-time trajectories, control system.

Key Terms. Mathematical Model, Specification Process, Verification Process

1 Introduction

A subset of the state space of a control system is called viable, if for any initial point in this set there exists a solution of the control system which stays forever in this set. Usual problems associated with viability are checking if a given set is viable, finding a solution (and/or the corresponding control input) which stays forever in this set (viable solution), designing a viable region [2]. Viability was studied in many works on the theory of differential equations and inclusions and the control theory [20, 5, 2, 3, 9, 19, 24, 21, 7, 10, 1, 16, 6]. The corresponding results can be straightforwardly applied to control and verification problems for hybrid (discrete-continuous) systems [11] and other models of cyber-physical systems [22, 4, 17, 23], assuming that viable sets are interpreted as safety regions. However, this interpretation suggests certain natural generalizations of the notion of viability. We propose and investigate one such generalization in this paper.

Let $n \geq 1$ be a natural number, I be a non-empty finite set, and $f_i : \mathbb{R} \to \mathbb{R}^n$, $i \in I$ be an indexed family of vector fields.

Let $T = [0, +\infty)$, $\mathcal{I}$ be the set of all functions from T to I which are piecewise-constant on each compact segment $[a, b] \subset T$, and $\|\cdot\|$ denote the Euclidean norm on $\mathbb{R}^n$. Consider a switched dynamical system [18] of the form

$$\dot{x}(t) = f_{\sigma(t)}(t, x(t)) \tag{1}$$

where, $\sigma \in \mathcal{I}$, $t \geq 0$.

Assume that for each $i \in I$:

1. f_i is continuous and bounded on $[0, +\infty) \times \mathbb{R}^n$;
2. there exists a number $L > 0$ such that $\|f_i(t, x_1) - f_i(t, x_2)\| \leq L \|x_1 - x_2\|$ for all $x_1, x_2 \in \mathbb{R}^n$, $t \in T$, and $i \in I$ (Lipschitz-continuity).

Under these conditions Caratheodory existence theorem [8] implies that for each $t_0 \in T$ and $x_0 \in \mathbb{R}^n$, and $\sigma \in \mathcal{I}$ the problem

$$\frac{d}{dt}x(t) = f_{\sigma(t)}(t, x(t)) \tag{2}$$

$$x(t_0) = x_0 \tag{3}$$

has a Caratheodory solution defined for all $t \geq t_0$, i.e. a function $t \mapsto x(t; t_0; x_0; u)$ which is absolutely continuous on every segment $[a, b] \subset [t_0, +\infty)$, satisfies the equation (2) a.e. (almost everywhere in the sense of Lebesgue measure), and satisfies (3). Moreover, this solution is unique in the sense that for any function $x : [t_0, t_1) \to \mathbb{R}^n$, which is absolutely continuous on every segment $[a, b] \subset [t_0, t_1)$, satisfies (2) a.e. on $[t_0, t_1)$ and satisfies (3), $x(t) = x(t; t_0; x_0; u)$ holds for $t \in [t_0, t_1)$.

For any $X \subseteq \mathbb{R}^n$ and $x_0 \in X$ denote by $VS(X, x_0)$ (set of viable switchings) the set of all $\sigma \in \mathcal{I}$ such that $x(t; 0; x_0; \sigma) \in X$ for all $t \geq 0$;

If $VS(X, x_0) \neq \emptyset$ for each $x_0 \in X$, then X is a viable set of (1) and functions $t \mapsto x(t; 0; x_0; \sigma)$, $\sigma \in VS(X, x_0)$ are viable solutions for X.

Let $Y \subseteq \mathbb{R}^n$ be a set. Let us say that a set $X \subseteq \mathbb{R}^n$ is *Y-strongly viable*, if for each $x_0 \in X$ there exists $\sigma \in VS(X, x_0)$ such that $\sigma(t)$ is constant on each interval $(t_1, t_2) \subset [0, +\infty)$ such that $x(t; 0; x_0; \sigma) \in Y$ for all $t \in (t_1, t_2)$.

In particular, X is viable if and only if X is $\emptyset$-strongly viable. Thus strong viability is a generalization of viability.

This notion has the following natural interpretation: the state of the system (1) can be forced to stay in a given "safe" set X by applying a specific control strategy (σ) only when its state is outside Y. When the state of the system is inside Y, one can keep the control constant (i.e. do not make any switchings) without the risk of driving the system into the "unsafe" region $\mathbb{R}^n \backslash X$. Then Y can be interpreted as a set of states where "nothing specific needs to be done" to ensure safety of the system and the complement of Y can be interpreted as a set of states upon reaching which "something may need to be done" to ensure safety.

In this paper we will consider the case when X is the complement of the origin (i.e. the origin may be interpreted as a safety hazard) and propose a convenient

sufficient condition which can be used to verify that for a given system, X, and Y, X is Y-strongly viable.

To do this we will use the notion of a Nondeterministic Complete Markovian System (NCMS) [14] which is based on the notion of a solution system by O. Hájek [12]. More specifically, we will represent the system (1) using a suitable NCMS and reduce the problem of Y-strong viability of a set X to the problem of the existence of global-in-time trajectories of NCMS which was investigated in [14, 15] and apply a theorem about the right dead-end path in NCMS [15] in order to obtain a condition of Y-strong viability.

To make the paper self-contained, in Section 2 we give the necessary definitions and facts about NCMS. In Section 3 we formulate and prove the main result of the paper.

2 Preliminaries

2.1 Notation

We will use the following notation: $\mathbb{N} = \{1, 2, 3, ...\}$, $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, $\mathbb{R}$ is the set of real numbers, $\mathbb{R}_+$ is the set of nonnegative real numbers, $f : A \to B$ is a total function from a set A to a set B, $f : A \tilde{\to} B$ denotes a partial function from a set A to a set B. We will denote by 2^A the power set of a set A and by $f|_A$ the restriction of a function f to a set A.

If A, B are sets, then B^A will denote the set of all total functions from A to B and ${}^A B$ will denote the set of all partial function from A to B.

For a function $f : A \tilde{\to} B$ the symbol $f(x) \downarrow$ ($f(x) \uparrow$) mean that $f(x)$ is defined, or, respectively, undefined on the argument x.

We will not distinguish the notions of a function and a functional binary relation. When we write that a function $f : A \tilde{\to} B$ is total or surjective, we mean that f is total on the set A specifically ($f(x)$ is defined for all $x \in A$), or, respectively, is onto B (for each $y \in B$ there exists $x \in A$ such that $y = f(x)$).

We will use the following notations for $f : A \tilde{\to} B$: $dom(f) = \{x \,|\, f(x) \downarrow\}$, i.e. the domain of f (note that in some fields like category theory the domain of a partial function is defined differently), and $range(f) = \{y \mid \exists x \; f(x) \downarrow \wedge \; y = f(x)\}$. We will use the same notation for the domain and range of a binary relation: if $R \subseteq A \times B$, then $dom(R) = \{x \,|\, \exists\, y \;\; (x, y) \in R\}$ and $range(R) = \{y \,|\, \exists\, x \;\; (x, y) \in R\}$.

We will denote by $f(x) \cong g(x)$ the strong equality (where f and g are partial functions): $f(x) \downarrow$ if and only if $g(x) \downarrow$, and $f(x) \downarrow$ implies $f(x) = g(x)$.

We will denote by $f \circ g$ the functional composition: $(f \circ g)(x) \cong f(g(x))$.

For any set X and a value y we will denote by $X \mapsto y$ a constant function defined on X which takes the value y.

Also, we will denote by T the non-negative real time scale $[0, +\infty)$ and assume that T is equipped with a topology induced by the standard topology on $\mathbb{R}$.

The symbols $\neg$, $\vee$, $\wedge$, $\Rightarrow$, $\Leftrightarrow$ will denote the logical operations of negation, disjunction, conjunction, implication, and equivalence respectively.

2.2 Nondeterministic Complete Markovian Systems (NCMS)

The notion of a NCMS was introduced in [13] for studying the relation between the existence of global and local trajectories of dynamical systems. It is close to the notion of a solution system by O. Hájek [12], however there are some differences between these two notions [14].

Denote by $\mathfrak{T}$ the set of all intervals (connected subsets) in T which have the cardinality greater than one.

Let Q be a set (a state space) and Tr be some set of functions of the form $s : A \to Q$, where $A \in \mathfrak{T}$. The elements of Tr will be called *(partial) trajectories.*

Definition 1. *([13, 14]) A set of trajectories Tr is closed under proper restrictions (CPR), if $s|_A \in Tr$ for each $s \in Tr$ and $A \in \mathfrak{T}$ such that $A \subseteq dom(s)$.*

Definition 2. *([13, 14])*

(1) A trajectory $s_1 \in Tr$ is a subtrajectory of $s_2 \in Tr$ (denoted as $s_1 \sqsubseteq s_2$), if $dom(s_1) \subseteq dom(s_2)$ and $s_1 = s_2|_{dom(s_1)}$.
(2) A trajectory $s_1 \in Tr$ is a proper subtrajectory of $s_2 \in Tr$ (denoted as $s_1 \sqsubset s_2$), if $s_1 \sqsubseteq s_2$ and $s_1 \neq s_2$.
(3) Trajectories $s_1, s_2 \in Tr$ are incomparable, if neither $s_1 \sqsubseteq s_2$, nor $s_2 \sqsubseteq s_1$.

The set $(Tr, \sqsubseteq)$ is a (possibly empty) partially ordered set.

Definition 3. *([13, 14]) A CPR set of trajectories Tr is*

(1) Markovian (Fig. 2), if for each $s_1, s_2 \in Tr$ and $t \in T$ such that $t = \sup dom(s_1) = \inf dom(s_2)$, $s_1(t) \downarrow$, $s_2(t) \downarrow$, and $s_1(t) = s_2(t)$, the following function s belongs to Tr:
$$s(t) = \begin{cases} s_1(t), & t \in dom(s_1) \\ s_2(t), & t \in dom(s_2) \end{cases}$$
(2) complete, if each non-empty chain in $(Tr, \sqsubseteq)$ has a supremum.

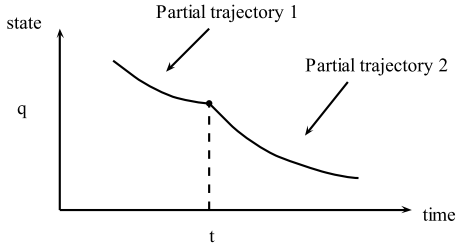


Fig. 1. Markovian property of NCMS. If one trajectory ends and another begins in the state q at time t, then their concatenation is a trajectory.

Definition 4. *([13, 14]) A nondeterministic complete Markovian system (NCMS) is a triple (T, Q, Tr), where Q is a set (state space) and Tr (trajectories) is a set of functions $s : T \tilde{\to} Q$ such that $dom(s) \in \mathfrak{T}$, which is CPR, complete, and Markovian.*

An overview of the class of all NCMS can be given using the notion of an *LR representation* [13–15].

Definition 5. *([13, 14]) Let $s_1, s_2 : T \tilde{\to} Q$. Then s_1 and s_2 coincide:*

(1) on a set $A \subseteq T$, if $s_1|_A = s_2|_A$ and $A \subseteq dom(s_1) \cap dom(s_2)$ (this is denoted as $s_1 \doteq_A s_2$);
(2) in a left neighborhood of $t \in T$, if $t > 0$ and there exists $t' \in [0, t)$ such that $s_1 \doteq_{(t',t]} s_2$ (this is denoted as $s_1 \doteq_{t-} s_2$);
(3) in a right neighborhood of $t \in T$, if there exists $t' > t$, such that $s_1 \doteq_{[t,t')} s_2$ (this is denoted as $s_1 \doteq_{t+} s_2$).

Let Q be a set. Denote by $ST(Q)$ the set of pairs (s, t) where $s : A \to Q$ for some $A \in \mathfrak{T}$ and $t \in A$.

Definition 6. *([13, 14]) A predicate $p : ST(Q) \to Bool$ is*

(1) left-local, if $p(s_1, t) \Leftrightarrow p(s_2, t)$ whenever $\{(s_1, t), (s_2, t)\} \subseteq ST(Q)$ and $s_1 \doteq_{t-} s_2$ hold, and, moreover, $p(s, t)$ holds whenever t is the least element of $dom(s)$;
(2) right-local, if $p(s_1, t) \Leftrightarrow p(s_2, t)$ whenever $\{(s_1, t), (s_2, t)\} \subseteq ST(Q)$ and $s_1 \doteq_{t+} s_2$ hold, and, moreover, $p(s, t)$ holds whenever t is the greatest element of $dom(s)$.

Let $LR(Q)$ be the set of all pairs (l, r), where $l : ST(Q) \to Bool$ is a left-local predicate and $r : ST(Q) \to Bool$ is a right-local predicate.

Definition 7. *([14]) A pair $(l, r) \in LR(Q)$ is called a LR representation of a NCMS $\Sigma = (T, Q, Tr)$, if*

$$Tr = \{s : A \to Q \mid A \in \mathfrak{T} \wedge (\forall t \in A\ l(s, t) \wedge r(s, t))\}.$$

The following theorem gives a representation of NCMS using predicate pairs.

Theorem 1. *([14, Theorem 1])*

(1) Each pair $(l, r) \in LR(Q)$ is a LR representation of a NCMS with the set of states Q.
(2) Each NCMS has a LR representation.

2.3 Existence global-in-time trajectories of NCMS

The problem of the existence of global trajectories of NCMS was considered in [13, 14] and was reduced to a more tractable problem of the existence of locally defined trajectories. Informally, the method of proving the existence of a global trajectory in NCMS consists of guessing a "region" (subset of trajectories) which presumably contains a global trajectory and has a convenient representation in the form of (another) NCMS and proving that this region indeed contains a global trajectory by finding or guessing certain locally defined trajectories independently in a neighborhood of each time moment.

Below we briefly state the main results about the existence of global trajectories of NCMS described in [15].

Let $\Sigma = (T, Q, Tr)$ be a fixed NCMS.

Definition 8. *([15]) Σ satisfies*

(1) local forward extensibility (LFE) property, if for each $s \in Tr$ of the form $s : [a, b] \to Q$ $(a < b)$ there exists a trajectory $s' : [a, b'] \to Q$ such that $s' \in Tr$, $s \sqsubseteq s'$ and $b' > b$.
(2) global forward extensibility (GFE) property, if for each trajectory s of the form $s : [a, b] \to Q$ there exists a trajectory $s' : [a, +\infty) \to Q$ such that $s \sqsubseteq s'$.

Definition 9. *([15]) A right dead-end path (in Σ) is a trajectory $s : [a, b) \to Q$, where $a, b \in T$, $a < b$, such that there is no $s' : [a, b] \to Q$, $s \in Tr$ such that $s \sqsubset s'$ (i.e. s cannot be extended to a trajectory on $[a, b]$).*

Definition 10. *([15]) An escape from a right dead-end path $s : [a, b) \to Q$ (in Σ) is a trajectory $s' : [c, d) \to Q$ (where $d \in T \cup \{+\infty\}$) or $s' : [c, d] \to Q$ (where $d \in T$) such that $c \in (a, b)$, $d > b$, and $s(c) = s'(c)$. An escape s' is called infinite, if $d = +\infty$.*

Definition 11. *([15]) A right dead-end path $s : [a, b) \to Q$ in Σ is called strongly escapable, if there exists an infinite escape from s.*

Definition 12. *([15])*
(1) A right extensibility measure is a function $f^+ : \mathbb{R} \times \mathbb{R} \tilde{\to} \mathbb{R}$ such that $A = \{(x, y) \in T \times T \mid x \leq y\} \subseteq dom(f^+)$, $f(x, y) \geq 0$ for all $(x, y) \in A$, $f^+|_A$ is strictly decreasing in the first argument and strictly increasing in the second argument, and for each $x \geq 0$, $f^+(x, x) = x$, $\lim_{y \to +\infty} f^+(x, y) = +\infty$.
(2) A right extensibility measure f^+ is called normal, if f^+ is continuous on $\{(x, y) \in T \times T \mid x \leq y\}$ and there exists a function α of class K_∞ (i.e. the function $\alpha : [0, +\infty) \to [0, +\infty)$ is continuous, strictly increasing, and $\alpha(0) = 0$, $\lim_{x \to +\infty} \alpha(x) = +\infty$) such that $\alpha(y) < y$ for all $y > 0$ and the function $y \mapsto f^+(\alpha(y), y)$ is of class K_∞.

An example of a right extensibility measure is $f_1^+(x, y) = 2y - x$.

Let f^+ be a right extensibility measure.

Definition 13. *([15]) A right dead-end path $s : [a, b) \to Q$ is called f^+-escapable, if there exists an escape $s' : [c, d] \to Q$ from s such that $d \geq f^+(c, b)$.*

Theorem 2. *([15], About right dead-end path) Assume that f^+ is a normal right extensibility measure and Σ satisfies LFE. Then each right dead-end path is strongly escapable if and only if each right dead-end path is f^+-escapable.*

Lemma 1. *([15]) Σ satisfies GFE if and only if Σ satisfies LFE and each right dead-end path is strongly escapable.*

Theorem 3. *([15], Criterion of the existence of global trajectories of NCMS)*
Let (l, r) be a LR representation of Σ. Then Σ has a global trajectory if and only if there exists a pair $(l', r') \in LR(Q)$ such that

(1) $l'(s, t) \Rightarrow l(s, t)$ and $r'(s, t) \Rightarrow r(s, t)$ for all $(s, t) \in ST(Q)$;
(2) $\forall t \in [0, \epsilon]$ $l'(s, t) \wedge r'(s, t)$ holds for some $\epsilon > 0$ and a function $s : [0, \epsilon] \to Q$;
(3) if (l', r') is a LR representation of a NCMS Σ', then Σ' satisfies GFE.

3 Main result

Let I, $\mathcal{I}$, and f_i, $i \in I$, and $x(t; t_0; x_0; \sigma)$ be defined as in Section 1. Let $X = \mathbb{R}^n \backslash \{0\}$ and $Y \subset \mathbb{R}^n$ be a set. Let denote $D = \mathbb{R}^n \backslash Y$.

Let us state the main result:

Theorem 4. *Assume that:*

(1) for each $t \in T$ there exist $i_1, i_2 \in I$ such that $f_{i_1}(t, 0)$ and $f_{i_2}(t, 0)$ are noncollinear;
(2) $\{0\}$ is a path-component of $\{0\} \cup Y$.

Then X is Y-strongly viable.

We will need several lemmas to prove this theorem.

Let us fix an element $x_0^* \in X$.

Let $Q = \mathbb{R}^n \times I$. Denote by $pr_1 : Q \to \mathbb{R}^n$, $pr_2 : Q \to I$ the projections on the first and second component, i.e. $pr_1((x_0, i)) = x_0$ and $pr_2((x_0, i)) = i$.

Let Tr be the set of all functions $s : A \to Q$, where $A \in \mathfrak{T}$, such that the following conditions are satisfied, where $x = pr_1 \circ s$ and $\sigma = pr_2 \circ s$:

1) σ is piecewise-constant on each segment $[a, b] \subseteq A$ $(a < b)$;

2) x is absolutely continuous on each segment $[a, b] \subseteq A$ $(a < b)$ and satisfies the equation $\frac{d}{dt}x(t) = f_i(t, x(t))$ a.e. on A;

3) $x(t) \neq 0$ for all $t \in A$;

4) for each non-maximal $t \in A$ such that $x(t) \notin D$ there exists $t' \in (t, +\infty) \cap A$ such that $\sigma(t'') = \sigma(t)$ for all $t'' \in [t, t')$;

5) for each non-minimal $t \in A$ such that $x(t) \notin D$ there exists $t' \in (0, t) \cap A$ such that $\sigma(t'') = \sigma(t)$ for all $t'' \in (t', t]$;

6) if $0 \in A$, then $x(0) = x_0^*$.

It follows straightforwardly from this definition that $\Sigma(x_0^*) = (T, Q, Tr)$ is a NCMS (i.e. Tr is a CPR, Markovian, and complete set of trajectories).

Let us find a sufficient condition which ensures that Σ has a global trajectory.

Lemma 2. *(1) $\Sigma(x_0^*)$ satisfies the LFE property.*
(2) There exists $s \in Tr$ and $\varepsilon > 0$ such that $dom(s) = [0, \varepsilon]$.

Proof. (1) Let $s : [a, b] \to Q$ be a trajectory, $x = pr_1 \circ s$, and $u = pr_2 \circ s$. Let $\sigma' : [a, +\infty) \to I$ be a function such that $\sigma'(t) = \sigma(t)$, if $t \in [a, b]$ and $\sigma'(t) = \sigma(b)$, if $t > b$. Then $\sigma = \sigma'|_{[a,b]}$, σ' is piecewise-constant on each segment in its domain, and $x(t) = x(t; a; x(a); \sigma')$ for all $t \in [a, b]$. Let $b' = b + 1$ and $x' : [a, b'] \to \mathbb{R}^n$ be a function such that $x'(t) = x(t; a; x(a); \sigma')$ for $t \in [a, b']$. Then $x = x'|_{[a,b]}$. Because $x'(t) \neq 0$ for all $t \in [a, b]$ and x' is continuous, there exists $b'' \in (b, b']$ such that $x'(t) \neq 0$ for all $t \in [a, b'']$. Let $s' : [a, b''] \to Q$ be a function such that $s'(t) = (x'(t), \sigma'(t))$ for all $t \in [a, b'']$. Then it follows immediately that $s' \in Tr$. Besides, $s \sqsubseteq s'$. Thus Σ satisfies LFE.

(2) Let us choose any $i_0 \in I$ and define $x : T \to \mathbb{R}^n$ as $x(t) = x(t; 0; x_0^*; \sigma_0)$ for all $t \in T$, where $\sigma_0(t) = i_0$ for all t. Then x is continuous and $x(0) = x_0^* \neq 0$, so there exists $\varepsilon > 0$ such that $x(t) \neq 0$ for all $t \in [0, \varepsilon]$. Let $s : [0, \varepsilon] \to Q$ be a function $s(t) = (x(t), i_0)$, $t \in [0, \varepsilon]$. Then $s \in Tr$. □

Lemma 3. *Assume that:*

(1) for each $t \in T$ there exist $i_1, i_2 \in I$ such that $f_{i_1}(t, 0)$, $f_{i_2}(t, 0)$ are (nonzero) noncollinear vectors, i.e. $k_1 f_{i_1}(t, 0) + k_2 f_{i_2}(t, 0) \neq 0$ whenever $k_1, k_2 \in \mathbb{R}$ are not both zero;
(2) for each $s \in Tr$ defined on a set of the form $[t_1, t_2)$, if $\lim_{t \to t_2 -} (pr_1 \circ s)(t) = 0$, then $pr_1(s(t)) \in D$ for some $t \in [t_1, t_2)$.

Then each right dead-end path in $\Sigma(x_0^*)$ is f_1^+-escapable, where $f_1^+(x, y) = 2y - x$ is a right extensibility measure.

Proof. Let $M' = 1 + \sup\{\|f_i(t', x')\| \,|(t', x') \in T \times \mathbb{R}^n, i \in I\}$. Then $0 < M' < +\infty$, because f is bounded.

Let $s : [a, b) \to Q$ be a right dead-end path and $x = pr_1 \circ s$, $\sigma = pr_2 \circ s$. Let $\sigma' : [a, +\infty) \to I$ be a function such that $\sigma'(t) = \sigma(t)$, if $t \in [a, b)$ and $\sigma'(t) = \sigma(a)$, if $t \geq b$. Then $\sigma = \sigma'|_{[a,b)}$, σ' is Lebesgue-measurable, and $x(t) = x(t; a; x(a); \sigma')$ for all $t \in [a, b)$. Then there exists a limit $x_l = \lim_{t \to b-} x(t) = x(b; a; x(a); \sigma') \in \mathbb{R}^n$.

Firstly, consider the case when $x_l \neq 0$. Then $\|x_l\| > 0$. Let us choose an arbitrary $t_0 \in (a, b)$ such that $b - t_0 < \|x_l\| / (4M')$ and $\|x(t_0) - x_l\| < \|x_l\| / 2$ (this is possible, because $x_l = \lim_{t \to b-} x(t)$). Let $\sigma'' : [t_0, +\infty) \to I$ and $x'' : [t_0, +\infty) \to \mathbb{R}^n$ be functions such that $\sigma''(t) = \sigma(t_0)$ for all $t \geq t_0$ and $x''(t) = x(t; t_0; x(t_0); \sigma'')$ for all $t \geq t_0$. Then $\|x''(t_0)\| = \|x(t_0) - x_l + x_l\| \geq \|x_l\| - \|x(t_0) - x_l\| > \|x_l\| / 2 > 2M'(b - t_0)$. Then for all $t \geq t_0$ we have

$$\|x''(t)\| = \left\| x''(t_0) + \int_{t_0}^{t} f_{\sigma''(t)}(t, x''(t)) dt \right\| \geq$$

$$\geq \|x''(t_0)\| - \int_{t_0}^{t} \left\| f_{\sigma''(t)}(t, x''(t)) \right\| dt >$$

$$> 2M'(b - t_0) - M'(t - t_0) = M'(2b - t_0 - t).$$

Let $d = 2b - t_0$. Then $d > t_0$ because $t_0 < b$. Then $x''(t) \neq 0$ for all $t \in [t_0, d]$. Let $s_* : [t_0, d] \to Q$ be a function such that $s_*(t) = (x''(t), \sigma''(t))$ for all $t \in [t_0, d]$. It follows immediately that $s_* \in Tr$. Also, $s_*(t_0) = s(t_0)$ and $d = 2b - t_0 = f_1^+(t_0, b)$. Then s_* is an escape from s and s is f_1^+-escapable.

Now consider the case when $x_l = 0$.

Let us choose $i_1, i_2 \in I$ such that $v_1 = f_{i_1}(b, 0)$ and $v_2 = f_{i_2}(b, 0)$ are noncollinear (this is possible by the assumption 1 of the lemma). Then the function $h(k_1, k_2) = \|k_1 v_1 + k_2 v_2\|$ attains some minimal value $M > 0$ on $\{(k_1, k_2) \in \mathbb{R} \times \mathbb{R} \mid |k_1| + |k_2| = 1\}$. Then for all k_1, k_2 such that $k_1 \neq 0$ or $k_2 \neq 0$,

$$h(k_1, k_2) = (|k_1| + |k_2|)h(k_1(|k_1| + |k_2|)^{-1}, k_2(|k_1| + |k_2|)^{-1}) \geq M(|k_1| + |k_2|).$$

Let $\varepsilon = M/2 > 0$. Because f is continuous, there exists $\delta > 0$ such that for each $j = 1, 2$, $t \in T$, and $x_0 \in \mathbb{R}^n$ such that $|b - t| + \|x_0\| < \delta$ we have $\left\|f_{i_j}(t, x_0) - v_j\right\| = \left\|f_{i_j}(t, x_0) - f_{i_j}(b, 0)\right\| < \varepsilon$. Let $R = \delta/4$, $t_1 = \max\{b - R, a\}$, and $t_2 = b + R$. Then $R > 0$, $a \leq t_1 < b < t_2$ and for all $j = 1, 2$, $t \in [t_1, t_2]$ and x_0 such that $\|x_0\| \leq R$, $\left\|f_{i_j}(t, x_0) - v_j\right\| < \varepsilon$.

Let us choose an arbitrary $c \in (t_1, b)$ such that $b - c < \min\{R/(2M'), R/2\}$. Then $s|_{[c,b)} \in Tr$ by the CPR property and $\lim_{t \to t_2 -}(pr_1 \circ s|_{[c,b)})(t) = x_l = 0$, so by the assumption 2 there exists $t_0 \in [c, b)$ such that $pr_1(s(t_0)) = x(t_0) \in D$.

Let $x_1 : [t_0, t_2] \to \mathbb{R}^n$ and $x_2 : [t_0, t_2] \to \mathbb{R}^n$ be functions such that $x_1(t) = x(t; t_0; x(t_0); \sigma_1)$ and $x_2(t) = x(t; t_0; x(t_0); \sigma_2)$ for all $t \in [t_0, t_2]$, where $\sigma_j(t) = i_j$ for all t. Denote $d_j(t) = f_{i_j}(t, x_j(t)) - v_j$ for each $j = 1, 2$ and $t \in [t_0, t_2]$.

Then the following two cases are possible.

a) There exists $j \in \{1, 2\}$ such that $0 \notin range(x_j)$. Let us choose any $d \in (\max\{2b - t_0, t_0\}, t_2)$ (this is possible, because $t_0 < b < t_2$ and $2b - t_0 \leq 2b - c < b + R/2 < b + R = t_2$). Then let $s_* : [t_0, d] \to Q$ be a function such that $s_*(t_0) = s(t_0) = (x(t_0), \sigma(t_0))$ and $s_*(t) = (x_j(t), i_j)$ for all $t \in (t_0, d]$. Because $x_j(t_0) = x(t_0) \in D$ and $x_j(t) \neq 0$ for all $t \in [t_0, t_2] \supset [t_0, d]$, we have that $s_* \in Tr$. Besides, $s_*(t_0) = s(t_0)$ and $d > 2b - t_0 = f_1^+(t_0, b)$, so s_* is an escape from s and s is f_1^+-escapable.

b) $0 \in range(x_1) \cap range(x_2)$. Then because x_1, x_2 are continuous, there exist $t'_j = \min\{t \in [t_0, t_2] \mid x_j(t) = 0\}$ for $j = 1, 2$. Moreover, $t'_j \in (t_0, t_2]$ for $j = 1, 2$, because $x_1(t_0) = x_2(t_0) = x(t_0) \neq 0$.

If we suppose that $\|x_j(t)\| < R$ for each $j = 1, 2$ and $t \in [t_0, t'_j]$, then $\|d_j(t)\| = \left\|f_{i_j}(t, x_j(t)) - v_j\right\| < \varepsilon$ for each $j = 1, 2$ and $t \in [t_0, t'_j]$, whence

$$\|0 - 0\| = \|x_1(t'_1) - x_2(t'_2)\| =$$

$$= \left\| x(t_0) + \int_{t_0}^{t'_1} f_{i_1}(t, x_1(t))dt - x(t_0) - \int_{t_0}^{t'_2} f_{i_2}(t, x_2(t))dt \right\| =$$

$$= \left\| \int_{t_0}^{t'_1} v_1 + d_1(t)dt - \int_{t_0}^{t'_2} v_2 + d_2(t)dt \right\| =$$

$$= \left\| v_1(t_1' - t_0) - v_2(t_2' - t_0) + \int_{t_0}^{t_1'} d_1(t)dt - \int_{t_0}^{t_2'} d_2(t)dt \right\| \geq$$

$$\geq \|v_1(t_1' - t_0) - v_2(t_2' - t_0)\| - \int_{t_0}^{t_1'} \|d_1(t)\|\, dt - \int_{t_0}^{t_2'} \|d_2(t)\|\, dt \geq$$

$$\geq M(|t_1' - t_0| + |t_2' - t_0|) - \varepsilon(t_1' - t_0) - \varepsilon(t_2' - t_0) = \frac{M}{2}(t_1' - t_0 + t_2' - t_0) > 0.$$

We have a contradiction, so there exists $j \in \{1, 2\}$ and $t'' \in [t_0, t_j']$ such that $\|x_j(t'')\| \geq R$. This implies that

$$R \leq \|x_j(t'')\| = \left\|x_j(t_j') - x_j(t'')\right\| = \left\| \int_{t''}^{t_j'} f_{i_j}(t, x_j(t))dt \right\| \leq M'(t_j' - t'').$$

Then $t_j' - t_0 \geq t_j' - t'' \geq R/M' > 2(b - c) \geq 2(b - t_0)$, so $t_j' > 2b - t_0$. Let us choose any $d \in (\max\{2b - t_0, t_0\}, t_j')$. Let $s_* : [t_0, d] \to Q$ be a function such that $s_*(t_0) = s(t_0) = (x(t_0), \sigma(t_0))$ and $s_*(t) = (x_j(t), i_j)$ for all $t \in (t_0, d]$. Because $x_j(t_0) = x(t_0) \in D$ and $x_j(t) \neq 0$ for all $t \in [t_0, t_j') \supset [t_0, d]$, we have $s_* \in Tr$. Besides, $s_*(t_0) = s(t_0)$ and $d > 2b - t_0 = f_1^+(t_0, b)$, so s_* is an escape from s and s is f_1^+-escapable. □

Lemma 4. *Assume that:*

(1) for each $t \in T$ there exist $i_1, i_2 \in I$ such that $f_{i_1}(t, 0)$ and $f_{i_2}(t, 0)$ are noncollinear;
(2) $\{0\}$ is a path-component of $\{0\} \cup Y$.

Then $\Sigma(x_0^*)$ has a global trajectory.

Proof. Let us show that the assumption 2 of Lemma 3 holds. Let $s \in Tr$, $dom(s) = [t_1, t_2)$ $(t_1 < t_2)$, $\lim_{t \to t_2 -}(pr_1 \circ s)(t) = 0$. Denote $x = pr_1 \circ s$. Suppose that $x(t) \notin D$ for all $t \in [t_1, t_2)$. Let $\gamma : [0, 1] \to \{0\} \cup (\mathbb{R}^n \backslash D)$ be a function such that $\gamma(\varepsilon) = x(t_1 + \varepsilon(t_2 - t_1))$, if $\varepsilon \in [0, 1)$ and $\gamma(1) = 0$. Then γ is continuous, so there is a path from $\gamma(0) = x(t_1) \neq 0$ to 0 in $\{0\} \cup (\mathbb{R}^n \backslash D) = \{0\} \cup Y$ (considered as a topological subspace of $\mathbb{R}^n$). This contradicts the assumption that $\{0\}$ is a path-component of $\{0\} \cup Y$. Thus $x(t) \in D$ for some $t \in [t_1, t_2)$.

The assumption 1 of Lemma 3 also holds, so by Lemma 2, Lemma 3, Lemma 1, Theorem 2, Σ satisfies GFE. Besides, by Lemma 2 there exists $s \in Tr$ with $dom(s) = [0, \varepsilon]$ for some $\varepsilon > 0$, so by the GFE property, Σ has a global trajectory. □

Proof (of Theorem 4). Follows straightforwardly from Lemma 4, because the statement of Lemma 4 holds for any $x_0^* \in X$.

4 Conclusion

We have proposed the notion of an Y-strongly viable set X for nonlinear switched systems. This notion follows naturally from interpretation of viable sets as safety regions. We have considered the case when X is the complement of the origin (i.e. the origin may be interpreted as a safety hazard) and proposed a convenient sufficient condition which can be used to verify that for a given system, X, and Y, X is Y-strongly viable. In the forthcoming papers we plan to investigate other cases give the corresponding conditions.

References

1. D. Angeli and E. D. Sontag. Forward completeness, unboundedness observability, and their lyapunov characterizations. *Systems & Control Letters*, 38(4):209–217, 1999.
2. J.-P. Aubin. *Viability Theory (Modern Birkhauser Classics)*. Birkhauser Boston, 2009.
3. J. P. Aubin and A. Cellina. *Differential inclusions: set-valued maps and viability theory*. Springer-Verlag GmbH, 1984.
4. R. Baheti and H. Gill. Cyber-physical systems. *The Impact of Control Technology*, pages 161–166, 2011.
5. J. Bebernes and J. Schuur. The wazewski topological method for contingent equations. *Annali di Matematica Pura ed Applicata*, 87(1):271–279, 1970.
6. O. Cârjă, M. Necula, and I. I. Vrabie. *Viability, invariance and applications*, volume 207. Elsevier Science Limited, 2007.
7. E. A. Coddington and N. Levinson. *Theory of Ordinary Differential Equations*. Krieger Pub Co, 1984.
8. A. Filippov. *Differential Equations with Discontinuous Righthand Sides: Control Systems (Mathematics and its Applications)*. Springer, 1988.
9. H. Frankowska and S. Plaskacz. A measurable upper semicontinuous viability theorem for tubes. *Nonlinear analysis*, 26(3):565–582, 1996.
10. Y. E. Gliklikh. Necessary and sufficient conditions for global-in-time existence of solutions of ordinary, stochastic, and parabolic differential equations. In *Abstract and Applied Analysis*, volume 2006, pages 1–17. MANCORP PUBLISHING, 2006.
11. R. Goebel, R. G. Sanfelice, and A. Teel. Hybrid dynamical systems. 29(2):28–93, 2009.
12. O. Hájek. Theory of processes, i. *Czechoslovak Mathematical Journal*, 17:159–199, 1967.
13. I. Ivanov. A criterion for existence of global-in-time trajectories of non-deterministic Markovian systems. *Communications in Computer and Information Science (CCIS)*, 347:111–130, 2013.
14. I. Ivanov. On existence of total input-output pairs of abstract time systems. *Communications in Computer and Information Science (CCIS)*, 412:308–331, 2013.
15. I. Ivanov. On representations of abstract systems with partial inputs and outputs. In T. Gopal, M. Agrawal, A. Li, and S. Cooper, editors, *Theory and Applications of Models of Computation*, volume 8402 of *Lecture Notes in Computer Science*, pages 104–123. Springer International Publishing, 2014.
16. G. Labinaz and M. Guay. *Viability of Hybrid Systems: A Controllability Operator Approach*. Springer Netherlands, 2012.

17. E. A. Lee and S. A. Seshia. *Introduction to embedded systems: A cyber-physical systems approach*. Lulu.com, 2013.
18. D. Liberzon. *Switching in Systems and Control (Systems & Control: Foundations & Applications)*. Birkhauser Boston Inc., 2003.
19. M. D. M. Marques. Viability results for nonautonomous differential inclusions. *Journal of Convex Analysis*, 7(2):437–443, 2000.
20. M. Nagumo. Über die Lage der Integralkurven gewöhnlicher Differentialgleichungen. 1942.
21. S. W. Seah. Existence of solutions and asymptotic equilibrium of multivalued differential systems. *Journal of Mathematical Analysis and Applications*, 89(2):648–663, 1982.
22. J. Shi, J. Wan, H. Yan, and H. Suo. A survey of cyber-physical systems. In *Wireless Communications and Signal Processing (WCSP), 2011 International Conference on*, pages 1–6. IEEE, 2011.
23. J. Sifakis. Rigorous design of cyber-physical systems. In *Embedded Computer Systems (SAMOS), 2012 International Conference on*, pages 319–319. IEEE, 2012.
24. I. I. Vrabie. A Nagumo type viability theorem. *An. Stiint. Univ. Al. I. Cuza Iasi. Mat.(NS)*, 51:293–308, 2005.

Natural Computing Modelling of the Polynomial Space Turing Machines

Bogdan Aman and Gabriel Ciobanu

Romanian Academy, Institute of Computer Science
Blvd. Carol I no.11, 700506 Iaşi, Romania
baman@iit.tuiasi.ro, gabriel@info.uaic.ro

Abstract. In this paper we consider a bio-inspired description of the polynomial space Turing machines. For this purpose we use membrane computing, a formalism inspired by the way living cells are working. We define and use logarithmic space systems with active membranes, employing a binary representation in order to encode the positions on the Turing machine tape.

Keywords. Natural computing, membrane computing, Turing machines.
Key Terms. MathematicalModel, Research.

1 Introduction

Membrane computing is a branch of the natural computing inspired by the architecture and behaviour of living cells. Various classes of membrane systems (also called P systems) have been defined in [12], together with their connections to other computational models. Membrane systems are characterized by three features: (i) a membrane structure consisting of a hierarchy of membranes (which are either disjoint or nested), with an unique top membrane called the *skin*; (ii) multisets of objects associated with membranes; (iii) rules for processing the objects and membranes. When membrane systems are seen as computing devices, two main research directions are usually considered: computational power in comparison with the classical notion of Turing computability (e.g., [2]), and efficiency in algorithmically solving NP-complete problems in polynomial time (e.g., [3]). Such efficient algorithms are obtained by trading space for time, with the space grown exponentially in a linear time by means of bio-inspired operations (e.g., membrane division). Thus, membrane systems define classes of computing devices which are both powerful and efficient.

Related to the investigations of these research directions, there have been studied several applications of these systems; among them, modelling of various biological phenomena and the complexity and emergent properties of such systems presented in [7]. In [4] it is presented the detailed functioning of the sodium-potassium pump, while in [1] it is described and analyzed the immune system in the formal framework of P systems.

In this paper we consider P systems with active membranes [11], and show that they provide an interesting simulation of polynomial space Turing machines by using only logarithmic space and a polynomial number of read-only input objects.

2 Preliminaries

We consider P systems with active membranes extended with an input alphabet, and such that the input objects cannot be created during the evolution [14]. The original definition also includes dissolution and division rules, rules that are not needed here. The version used in this paper is similar to evolution-communication P systems used in [6] with additional read-only input objects and polarities.

Definition 1. *A* P system with active membranes and input objects *is a tuple* $\Pi = (\Gamma, \Delta, \Lambda, \mu; w_1, \ldots, w_d, R)$*, where:*

- $d \geq 1$ *is the* initial degree*;*
- Γ *is a finite non-empty alphabet of* objects*;*
- Δ *is an* input alphabet *of objects such that* $\Delta \cap \Gamma = \emptyset$*;*
- Λ *is a finite* set of labels *for membranes;*
- μ *is a* membrane structure *(i.e., a rooted unordered tree, usually represented by nested brackets) in which each membrane is labelled by an element of* Λ *in a one-to-one way, and possesses an attribute called* electrical charge*, which can be either neutral (0), positive (+) or negative (-);*
- $w_1, \ldots, w_d$ *are strings over* Γ*, describing the* initial multisets *of objects placed in a number of* d *membranes of* μ*; notice that* w_i *is assigned to membrane* i*;*
- R *is a finite* set of rules *over* $\Gamma \cup \Delta$*:*
  1. $[a \to w]_h^\alpha$ object evolution rules
  An object $a \in \Gamma$ *is rewritten into the multiset* w*, if* a *is placed inside a membrane labelled by* h *with charge* α*. An object* a *can be deleted by considering* w *the empty multiset* $\emptyset$*. Notice that these rules allow only to rewritten objects from* Γ*, but not from* Δ*.*
  2. $a[\ ]_h^\alpha \to [b]_h^\beta$ send-in communication rules
  An object a *is sent into a membrane labelled by* h *and with charge* α*, becoming* b*; also, the charge of* h *is changed to* β*. If* $b \in \Delta$*, then* $a = b$ *must hold.*
  3. $[a]_h^\alpha \to b[\ ]_h^\beta$ send-out communication rules
  An object a*, placed into a membrane labelled by* h *and having charge* α*, is sent out of membrane* h *and becomes* b*; simultaneously, the charge of* h *is changed to* β*. If* $b \in \Delta$*, then* $a = b$ *must hold.*

Each configuration $\mathcal{C}_i$ of a P system with active membranes and input objects is described by the current membrane structure, including the electrical charges, together with the multisets of objects located in the corresponding membranes. The initial configuration of such a system is denoted by $\mathcal{C}_0$. An evolution step from the current configuration $\mathcal{C}_i$ to a new configuration $\mathcal{C}_{i+1}$, denoted by $\mathcal{C}_i \Rightarrow \mathcal{C}_{i+1}$, is done according to the principles:

- Each object and membrane is involved in at most one communication rule per step.
- Each membrane could be involved in several object evolution rules that can be applied in parallel inside it.
- The application of rules is maximally parallel: the only objects and membranes that do not evolve are those associated with no rule, or only to rules that are not applicable due to the electrical charges.
- When several conflicting rules could be applied at the same time, a nondeterministic choice is performed; this implies that multiple configurations can be reached as the result of an evolution step.
- In each computation step, all the chosen rules are applied simultaneously.
- Any object sent out from the skin membrane cannot re-enter it.

A *halting evolution* of such a system Π is a finite sequence of configurations $\overrightarrow{\mathcal{C}} = (\mathcal{C}_0, \ldots, \mathcal{C}_k)$, such that $\mathcal{C}_0 \Rightarrow \mathcal{C}_1 \Rightarrow \ldots \Rightarrow \mathcal{C}_k$, and no rules can be applied any more in $\mathcal{C}_k$. A *non-halting* evolution $\overrightarrow{\mathcal{C}} = (\mathcal{C}_i \mid i \in \mathbb{N})$ consists of an infinite evolution $\mathcal{C}_0 \Rightarrow \mathcal{C}_1 \Rightarrow \ldots$, where the applicable rules are never exhausted.

Example 1. Addition is trivial; we consider n objects a and m objects b placed in a membrane 0 with charge $+$. The rule $[b \rightarrow a]_+$ says that an object b is transformed in one object a. Such a rule is applied in parallel as many times as possible. Consequently, all objects b are erased. The remaining number of objects a represents the addition $n + m$. More examples can be found in [5].

In order to solve decision problems (i.e., decide languages over an alphabet Σ), we use *families* of recognizer P systems $\mathbf{\Pi} = \{\Pi_x \mid x \in \Sigma^*\}$ that respect the following conditions: (1) all evolutions halt; (2) two additional objects *yes* (successful evolution) and *no* (unsuccessful evolution) are used; (3) one of the objects *yes* and *no* appears in the halting configuration [13]. Each input x is associated with a P system Π_x that decides the membership of x in the language $L \subseteq \Sigma^*$ by accepting or rejecting it. The mapping $x \vdash \Pi_x$ must be efficiently computable for each input length [10].

In this paper we use a logarithmic space uniformity condition [14].

Definition 2. *A family of P systems $\mathbf{\Pi} = \{\Pi_x \mid x \in \Sigma^*\}$ is said to be $(\mathbf{L}, \mathbf{L})$-uniform if the mapping $x \vdash \Pi_x$ can be computed by two deterministic logarithmic space Turing machines F (for "family") and E (for "encoding") as follows:*

- *F computes the mapping $1^n \vdash \Pi_n$, where Π_n represents the membrane structure with some initial multisets and a specific input membrane, while n is the length of the input x.*
- *E computes the mapping $x \vdash w_x$, where w_x is a multiset encoding the specific input x.*
- *Finally, Π_x is Π_n with w_x added to the multiset placed inside its input membrane.*

In the following definition of space complexity adapted from [14], the input objects do not contribute to the size of the configuration of a P system. In this way, only the actual working space of the P system is measured, and P systems working in sublinear space may be analyzed.

Definition 3. *Given a configuration $\mathcal{C}$, the space size $|\mathcal{C}|$ is defined as the sum of the number of membranes in μ and the number of objects in Γ it contains. If $\overrightarrow{\mathcal{C}}$ is a halting evolution of Π, then $|\overrightarrow{\mathcal{C}}| = max\{|\mathcal{C}_0|, \ldots, |\mathcal{C}_k|\}$ or, in the case of a non-halting evolution $\overrightarrow{\mathcal{C}}$, $|\overrightarrow{\mathcal{C}}| = sup\{|\mathcal{C}_i| \mid i \in \mathbb{N}\}$. The space required by Π itself is then $|\Pi| = sup\{|\overrightarrow{\mathcal{C}}| \mid \overrightarrow{\mathcal{C}}$ is an evolution of $\Pi\}$.*

Notice that $|\Pi| = \infty$ if Π has an evolution requiring infinite space or an infinite number of halting evolutions that can occur such that for each $k \in \mathbb{N}$ there exists at least one evolution requiring most than k steps..

3 A Membrane Structure for Simulation

Let M be a single-tape deterministic Turing machine working in polynomial space $s(n)$. Let Q be the set of states of M, including the initial state s; we denote by $\Sigma' = \Sigma \cup \{\sqcup\}$ the tape alphabet which includes the blank symbol $\sqcup \notin \Sigma$. A computation step is performed by using $\delta : Q \times \Sigma' \to Q \times \Sigma' \times \{-1, 0, 1\}$, a (partial) transition function of M which we assume to be undefined on (q, σ) if and only if q is a final state. We describe a uniform family of P systems $\mathbf{\Pi} = \{\Pi_x \mid x \in \Sigma^*\}$ simulating M in logarithmic space.

Let $x \in \Sigma^n$ be an input string, and let $m = \lceil log\ s(n) \rceil$ be the minimum number of bits needed in order to write the tape cell indices $0, \ldots, s(n)$-1 in binary notation. The P system Π_n associated with the input length n and computed as $F(1^n)$ has a membrane structure consisting of $|\Sigma'| \cdot (m + 1) + 2$ membranes. The membrane structure contains:

- a skin membrane h;
- an inner membrane c (the input membrane) used to identify the symbol needed to compute the δ function;
- for each symbol $\sigma \in \Sigma'$ of M, the following set of membranes, linearly nested inside c and listed inward:
  - a membrane σ for each symbol σ of the tape alphabet Σ' of M;
  - for each $j \in \{0, \ldots, (m-1)\}$, a membrane labelled by j.

This labelling is used in order to simplify the notations. To respect the one-to-one labelling from Definition 1, the membrane j can be labelled j_σ. Thus in all rules using membranes j, the σ symbol is implicitly considered. Furthermore, the object z_0 is located inside the skin membrane h.

The encoding of x, computed as $E(x)$, consists of a set of objects describing the tape of M in its initial configuration on input x. These objects are the symbols of x subscripted by their position $bin(0), \ldots, bin(n-1)$ (where $bin(i)$ is the binary representation of i on m positions) in x, together with the $s(n) - n$

blank objects subscripted by their position $bin(n), \ldots, bin(s(n)-1)$. The binary representation, together with the polarities of the membranes, is essential when the membrane system has to identify the symbol needed to simulate the δ function (e.g., rule (13)). The multiset $E(x)$ is placed inside the input membrane c. Figure 1 depicts an example.

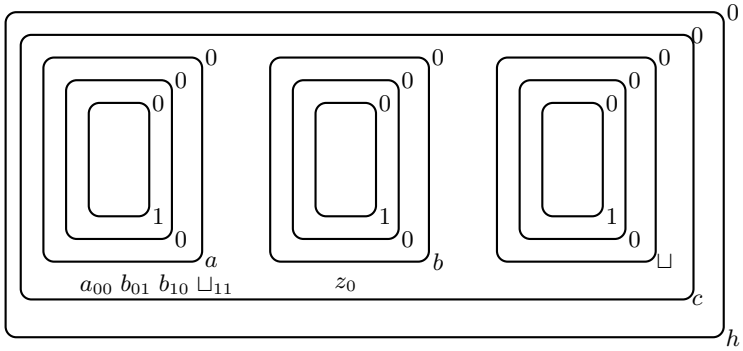


Fig. 1. Initial configuration of the P system Π_3 with tape alphabet $\Sigma' = \{a, b, ⊔\}$, working in space $s(n) = n + 1 = 4$ on the input abb.

During the first evolution steps of Π_x, each input object σ_i is moved from the input membrane c to the innermost membrane $(m-1)$ of the corresponding membrane σ by means of the following communication rules:

$$\sigma_i[\ ]^0_\sigma \to [\sigma_i]^0_\sigma \qquad \text{for } \ \sigma \in \Sigma',\ bin(0) \le i < bin(s(n)) \tag{1}$$

$$\sigma_i[\ ]^0_j \to [\sigma_i]^0_j \qquad \text{for } \ \sigma \in \Sigma',\ bin(0) \le i < bin(s(n)),\ 0 \le j < m \tag{2}$$

Since only one communication rule per membrane can be applied during each evolution step of Π_x, all $s(n)$ input objects pass through m membranes, in order to reach the innermost membranes $(m-1)$, in at most $l = s(n) + m$ evolution steps. In the meantime, the subscript of object z_0 is incremented up to $max\{0, l-3\}$ before object z_{l-3} exits and enters membrane c changing the membrane charge from 0 to +:

$$[z_t \to z_{t+1}]^0_c \qquad \text{for } \ 0 \le t < l-3 \tag{3}$$

$$[z_{l-3}]^0_c \to z_{l-3}[\ ]^0_c \tag{4}$$

$$z_{l-3}[\ ]^0_c \to [z_{l-3}]^+_c \tag{5}$$

The object z_{l-3} is rewritten to a multiset of objects containing an object z' (used in rule (9)) and $|\Sigma'|$ objects z_+ (used in rules (7)

$$[z_{l-3} \to z' \underbrace{z_+ \cdots z_+}_{|\Sigma'| \text{ copies}}\ ]^+_c \tag{6}$$

The objects z_+ are used to change the charges from 0 to + for all membranes $\sigma \in \Sigma'$ using parallel communication rules, and then are deleted:

$$z_+[\ ]^0_\sigma \to [\#]^+_\sigma \qquad \text{for } \sigma \in \Sigma' \tag{7}$$

$$[\# \to \emptyset]^+_\sigma \qquad \text{for } \sigma \in \Sigma' \tag{8}$$

In the meantime, the object z' is rewritten into z'' (in parallel with rule (7)), and then, in parallel with rule (8), into s_{00} (where s is the initial state of M):

$$[z' \to z'']^+_c \tag{9}$$

$$[z'' \to s_{00}]^+_c \tag{10}$$

The configuration reached by Π_x encodes the initial configuration of M:

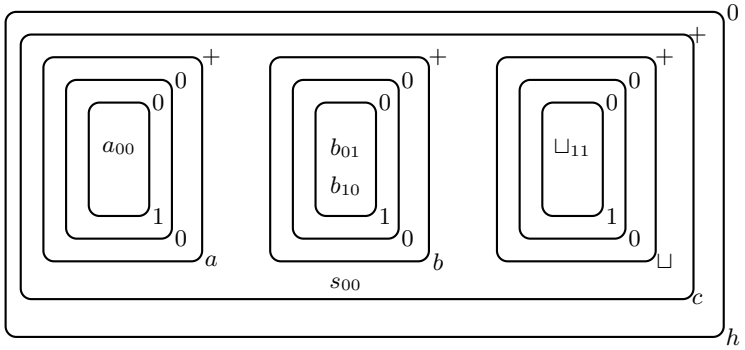


Fig. 2. Configuration of Π_x (from Figure 1) encoding the initial configuration of M on input $x = abb$ and using $s(|x|) = 4$ tape cells.

An arbitrary configuration of M on input x is encoded by a configuration of Π_x as it is described in Figure 3:

- membrane c contains the state-object q_i, where q is the current state of M and $i \in \{bin(0), \ldots, bin(s(n) - 1)\}$ is the current position of the tape head;
- membranes $(m - 1)$ contain all input objects;
- all other membranes are empty;
- all membranes are neutrally charged, except those labelled by $\sigma \in \Sigma'$ and by c which all are positively charged.

We employ this encoding because the input objects must be all located in the input membrane in the initial configuration of Π_x (hence they must encode both symbol and position on the tape), and they can never be rewritten.

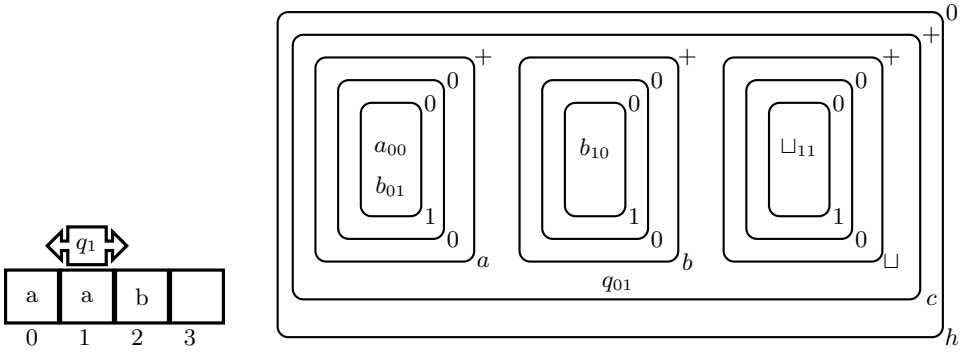


Fig. 3. A configuration of M (from Figure 1) and the corresponding configuration of Π_x simulating it. The presence of b_{01} inside membrane 1 of a indicates that tape cell 1 of M contains the symbol a.

4 Simulating Polynomial Space Turing Machines

Starting from a configuration of the single-tape deterministic Turing machine M, the simulation of a computation step of M by the membrane system Π_x is directed by the state-object q_i. As stated above, q_i encodes the current state of M and the position of the head on the tape (in binary format). To simulate the transition function δ of the Turing machine M in state q, it is necessary to identify the actual symbol occurring at tape position i. In order to identify this σ_i object from one of the $(m-1)$ membranes, the object q_i is rewritten into $|\Sigma'|$ copies of q_i', one for each membrane $\sigma \in \Sigma'$:

$$[q_i \to \underbrace{q_i' \cdots q_i'}_{|\Sigma'| \text{ copies}}]_c^+ \qquad q \in Q,\ bin(0) \le i < bin(s(n)) \tag{11}$$

The objects q_i' first enter the symbol-membranes in parallel, without changing the charges:

$$q_i'[\ ]_\sigma^+ \to [q_i']_\sigma^+ \quad \text{for} \quad \sigma \in \Sigma',\ q \in Q,\ bin(0) \le i < bin(s(n)) \tag{12}$$

The object q_i' traverses the membranes $0, \ldots, (m-1)$ while changing their charges such that they represent the bits of i from the least to the most significant one, where a positive charge is interpreted as 1 and a negative charge as 0. For instance, the charges of $[[[\ ]_2^-]_1^-]_0^+$ encode the binary number 001 (that is, decimal 1). By the j-th bit of a binary number is understood the bit from the j-th position when the number is read from right to left (e.g, the 0-th bit of the binary number 001 is 1). The changes of charges are accomplished by the rules:

$$q_i'[\ ]_j^0 \to [q_i']_j^\alpha \quad \text{for} \quad \sigma \in \Sigma',\ q \in Q,\ bin(0) \le i < bin(s(n)),\ 0 \le j < m, \tag{13}$$

where α is $-$ if the j-th bit of i is 0, and α is $+$ if the j-th bit of i is 1.

The membranes j, where $0 \le j < m$, behave now as "filters" for the input objects σ_k occurring in membrane $(m-1)$: these objects are sent out from each membrane j if and only if the j-th bit of k corresponds to the charge of j.

$$[\sigma_k]_j^\alpha \to [\ ]_j^\alpha \sigma_k \quad \text{for} \quad \sigma \in \Sigma',\ bin(0) \le k < bin(s(n)),\ 0 < j < m, \tag{14}$$

where α is $-$ if the j-th bit of k is 0, and α is $+$ if j-th bit of k is 1.

If an object σ_k reaches membrane 0, it is sent outside if the 0-th bit of k corresponds to the charge of membrane 0. In order to signal that it is the symbol occurring at location i of the tape, the charge of the corresponding membrane 0 is changed (either from $+$ to $-$ or from $-$ to $+$). By applying the rules (15) to (17), exactly one object σ_k, with $k = i$, will exit through membrane c:

$$\begin{array}{l} [\sigma_k]_0^+ \rightarrow [\ ]_0^- \sigma_k \quad \text{for} \ \ \sigma \in \Sigma', \, bin(0) \leq k < bin(s(n)), \\ [\sigma_k]_0^- \rightarrow [\ ]_0^+ \sigma_k \quad \text{for} \ \ \sigma \in \Sigma', \, bin(0) \leq k < bin(s(n)), \end{array} \tag{15}$$

where α is $-$ if the j-th bit of k is 0, and α is $+$ if the j-th bit of k is 1;

$$[\sigma_k]_\tau^+ \rightarrow \sigma_k [\ ]_\tau^- \quad \text{for} \ \ \sigma, \tau \in \Sigma', \, bin(0) \leq k < bin(s(n)). \tag{16}$$

After an σ_i exits from membrane c it gets blocked inside membrane h, by the new charge of membrane c, until it is allowed to move to its new location according to function δ of the Turing machine M. Thus, if another object τ_j reached membrane σ due to the new charge of membrane 0 established by rule (15), τ_j is contained in membrane σ until reintroduced in a membrane (m-1) using rule (2).

$$[\sigma_k]_c^+ \rightarrow \sigma_k [\ ]_c^- \quad \text{for} \ \ \sigma \in \Sigma', \, bin(0) \leq k < bin(s(n)) \tag{17}$$

Since there are $s(n)$ input objects, and each of them must traverse at most $(m + 1)$ membranes, the object σ_i reaches the skin membrane h after at most $l + 1$ steps, where l is as defined in Section 3 before rule (3). While the input objects are "filtered out", the state-object q_i' "waits" for l steps using the rules:

$$[q_i' \rightarrow q_{i,1}'']_{m-1}^\alpha \quad \text{for} \ \ \sigma \in \Sigma', \, q \in Q, \, bin(0) \leq i < bin(s(n)), \, \alpha \in \{-, +\} \tag{18}$$

$$\begin{array}{c} [q_{i,t}'' \rightarrow q_{i,t+1}'']_{m-1}^\alpha \quad \text{for} \ \ \sigma \in \Sigma', \, q \in Q, \, bin(0) \leq i < bin(s(n)), \\ \alpha \in \{-, +\}, \, 1 \leq t \leq l \end{array} \tag{19}$$

$$[q_{i,l+1}'' \rightarrow q_i'']_{m-1}^\alpha \quad \text{for} \ \ \sigma \in \Sigma', \, q \in Q, \, bin(0) \leq i < bin(s(n)), \, \alpha \in \{-, +\} \tag{20}$$

In order to reach membrane c, the objects q_i'' are sent out through membranes j $(0 < j \leq m - 1)$ using rule (21), through membrane 0 by rules (22) and (24), and through membranes $\sigma \in \Sigma'$ by rule (23). While passing through all these membranes, the charges are changed to neutral. This allows the input objects to move back to the innermost membrane $(m - 1)$ by using rules of type (2).

$$\begin{array}{c} [q_i'']_j^\alpha \rightarrow [\ ]_j^0 q_i'' \quad \text{for} \ \ \sigma \in \Sigma', \, q \in Q, \, bin(0) \leq i < bin(s(n)) \\ 0 < j \leq m - 1, \, \alpha \in \{-, +\} \end{array} \tag{21}$$

When q_i'' reaches the membranes 0, only one has the charge different from the 0-th bit of i, thus allowing q_i'' to identify the symbol in tape location i of M:

$$[q_i'']_0^\alpha \rightarrow [\ ]_0^0 q_i''' \quad \text{for} \ \ \sigma \in \Sigma', \, q \in Q, \, bin(0) \leq i < bin(s(n)), \tag{22}$$

where α is $-$ if the 0-th bit of i is 1, and α is $+$ if the 0-th bit of i is 0.

$$[q_i'''\,]_\sigma^- \to [\ ]_\sigma^0 q_{i,\sigma,1} \quad \text{for} \quad \sigma \in \Sigma',\, q \in Q,\, bin(0) \le i < bin(s(n)) \tag{23}$$

The other copies of q_i'' are sent out as objects # through membrane 0, and then deleted by rules of type (8):

$$[q_i''\,]_0^\alpha \to [\ ]_0^0 \# \quad \text{for} \quad \sigma \in \Sigma',\, q \in Q,\, bin(0) \le i < bin(s(n)), \tag{24}$$

where α is $-$ if the 0-th bit of i is 0, and α is $+$ if the 0-th bit of i is 1.

The state-object $q_{i,\sigma,1}$ waits in membrane c for l steps, l representing an upper bound of the number of steps needed for all the input objects to reach the innermost membranes:

$$[q_{i,\sigma,t} \to q_{i,\sigma,t+1}]_c^- \quad \text{for} \quad \sigma \in \Sigma',\, q \in Q,\, bin(0) \le i < bin(s(n)),\, 1 \le t < l \tag{25}$$

$$q_{i,\sigma,l}[\ ]_\sigma^0 \to [q_{i,\sigma,l}]_\sigma^+ \quad \text{for} \quad \sigma \in \Sigma',\, q \in Q,\, bin(0) \le i < bin(s(n)) \tag{26}$$

$$[q_{i,\sigma,l}]_\sigma^+ \to q'_{i,\sigma}[\ ]_\sigma^+ \quad \text{for} \quad \sigma \in \Sigma',\, q \in Q,\, bin(0) \le i < bin(s(n)) \tag{27}$$

The state-object $q'_{i,\sigma}$ now contains all the information needed to compute the transition function δ of the Turing machine M. Suppose $\delta(q,\sigma) = (r,v,d)$ for some $d \in \{-1,0,+1\}$. Then $q'_{i,\sigma}$ sets the charge of membrane v to $-$ and waits for $m+1$ steps, thus allowing σ_i to move to membrane $(m-1)$ of v by using the rules (31), (32) and (2):

$$q'_{i,\sigma}[\ ]_v^+ \to [q'_{i,\sigma}]_v^+ \quad \text{for} \quad \sigma, v \in \Sigma',\, q \in Q,\, bin(0) \le i < bin(s(n)) \tag{28}$$

$$[q'_{i,\sigma}]_v^+ \to q'_{i,\sigma,1}[\ ]_v^- \quad \text{for} \quad \sigma \in \Sigma',\, q \in Q,\, bin(0) \le i < bin(s(n)) \tag{29}$$

$$[q'_{i,\sigma,1}]_c^- \to q'_{i,\sigma,1}[\ ]_c^0 \quad \text{for} \quad \sigma \in \Sigma',\, q \in Q,\, bin(0) \le i < bin(s(n)) \tag{30}$$

$$\sigma_i[\ ]_c^0 \to [\sigma_i]_c^0 \quad \text{for} \quad \sigma \in \Sigma',\, bin(0) \le i < bin(s(n)) \tag{31}$$

$$\sigma_i[\ ]_v^- \to [\sigma_i]_v^- \quad \text{for} \quad \sigma, v \in \Sigma',\, q \in Q,\, bin(0) \le i < bin(s(n)) \tag{32}$$

$$[q'_{i,\sigma,t} \to q'_{i,\sigma,t+1}]_h^0 \quad \text{for} \quad \sigma \in \Sigma',\, q \in Q,\, bin(0) \le i < bin(s(n)),\ 1 \le t \le m \tag{33}$$

The object $q'_{i,\sigma,m+1}$ is used to change the charges of membranes c and v to $+$, thus preparing the system for the next step of the simulation:

$$q'_{i,\sigma,m+1}[\ ]_c^0 \to [q'_{i,\sigma,m+1}]_c^+ \quad \text{for} \quad \sigma \in \Sigma',\, q \in Q,\, bin(0) \le i < bin(s(n)) \tag{34}$$

$$q'_{i,\sigma,m+1}[\ ]_v^- \to [q'_{i,\sigma,m+1}]_v^- \quad \text{for} \quad \sigma, v \in \Sigma',\, q \in Q,\, bin(0) \le i < bin(s(n)) \tag{35}$$

$$[q'_{i,\sigma,m+1}]_v^- \to q''_{i,\sigma}[\ ]_v^+ \quad \text{for} \quad \sigma, v \in \Sigma',\, q \in Q,\, bin(0) \le i < bin(s(n)) \tag{36}$$

Finally, the state-object $q''_{i,\sigma}$ is rewritten to reflect the change of state and head position, thus producing a configuration of Π_x corresponding to the new configuration of M, as described in Section 3:

$$[q''_{i,\sigma} \to r_{i+d}]_c^+ \quad \text{for} \quad bin(0) \le i < bin(s(n)) \tag{37}$$

The P system Π_x is now ready to simulate the next step of M. If $q \in Q$ is a final state of M, we assume that $\delta(q, \sigma)$ is undefined for all $\sigma \in \Sigma'$; thus we introduce the following rules which halt the P system with the same result (acceptance or rejection) as M:

$$[q_i]_c^+ \rightarrow [\ ]_c^+ yes \quad \text{for} \quad bin(0) \leq i < bin(s(n)), \text{ if } q \text{ is an accepting state} \tag{38}$$

$$[yes]_h^0 \rightarrow [\ ]_h^0 yes \quad \text{for} \quad bin(0) \leq i < bin(s(n)), \text{ if } q \text{ is an accepting state} \tag{39}$$

$$[q_i]_c^+ \rightarrow [\ ]_c^+ no \quad \text{for} \quad bin(0) \leq i < bin(s(n)), \text{ if } q \text{ is a rejecting state} \tag{40}$$

$$[no]_h^0 \rightarrow [\ ]_h^0 no \quad \text{for} \quad bin(0) \leq i < bin(s(n)), \text{ if } q \text{ is a rejecting state} \tag{41}$$

The simulation directly leads to the following result.

Theorem 1. *Let M be a single-tape deterministic Turing machine working in polynomial space $s(n)$ and time $t(n)$. Then there exists an $(\mathbf{L}, \mathbf{L})$-uniform family $\mathbf{\Pi}$ of P systems Π_x with active membranes using object evolution and communication rules that simulates M in space $O(log\ n)$ and time $O(t(n)s(n))$.*

Proof. For each $x \in \Sigma^n$, the P system Π_x can be built from 1^n and x in logarithmic space as it is described in Definition 2; thus, the family $\mathbf{\Pi}$ is $(\mathbf{L}, \mathbf{L})$-uniform. Each P system Π_x uses only a logarithmic number of membranes and a constant number of objects per configuration; thus, Π_x works in space $O(log\ n)$. Simulating one of the $t(n)$ steps of M requires $O(s(n))$ time, an upper bound to the subscripts of objects used to introduce delays during the simulation; thus, the total time is $O(t(n)s(n))$.

5 Conclusion

In this paper we provided a simulation of the polynomial space Turing machines by using logarithmic space P systems with active membranes and binary representations for the positions on the tape. A similar approach is presented in [9]. There are important differences in terms of technical details and efficient representation; in comparison to [9], we improve the simulation by reducing the number of membranes (by $|\Sigma'| - 1$) and the number of rules (by $|\Sigma'| \cdot |Q| \cdot s(n) \cdot (5 - |\Sigma'|) + |\Sigma'| \cdot |\Sigma'| s(n) \cdot (2 \cdot m + 1) + |Q| \cdot s(n) - |\Sigma'| \cdot s(n) \cdot (m + 3)$). In particular, for the running example, the number of rules is reduced by $14 \cdot |Q| + 84$. A different approach is presented in [8] where it is claimed that a constant space is sufficient. However, in order to obtain the constant space space, input objects (from Δ) are allowed to create other objects (from Γ) leading to a different and more powerful formalism than the one used by us in this paper, and making such an approach not so interesting because of these unrealistic (powerful) input objects.

References

1. B. Aman, G.Ciobanu. Describing the Immune System Using Enhanced Mobile Membranes. *Electronic Notes in Theoretical Computer Science* **194**, 5–18 (2008).
2. B. Aman, G.Ciobanu. Turing Completeness Using Three Mobile Membranes. *Lecture Notes in Computer Science* **5715**, 42–55 (2009).
3. B. Aman, G. Ciobanu. *Mobility in Process Calculi and Natural Computing.* Natural Computing Series, Springer (2011).
4. D. Besozzi, G. Ciobanu. A P System Description of the Sodium-Potassium Pump. *Lecture Notes in Computer Science* **3365**, 210–223 (2004).
5. C. Bonchiş, G. Ciobanu, C. Izbaşa. Encodings and Arithmetic Operations in Membrane Computing. *Lecture Notes in Computer Science Volume* **3959**, 621–630 (2006).
6. M. Cavaliere. Evolution-Communication P Systems. *Lecture Notes in Computer Science* **2597**, 134–145 (2003).
7. G. Ciobanu, Gh. Păun, M.J. Pérez-Jiménez (Eds.). *Applications of Membrane Computing*. Springer (2006).
8. A. Leporati, L. Manzoni, G. Mauri, A.E. Porreca, C. Zandron. Constant-Space P Systems with Active Membranes. *Fundamenta Informaticae* **134**(1–2), 111–128 (2014).
9. A. Leporati, G. Mauri, A.E. Porreca, C. Zandron. A Gap in the Space Hierarchy of P Systems With Active Membranes. *Journal of Automata, Languages and Combinatorics* **19** (1-4), 173–184 (2014).
10. N. Murphy, D. Woods. The Computational Power of Membrane Systems Under Tight Uniformity Conditions. *Natural Computing* **10**, 613-632 (2011).
11. Gh. Păun. P Systems With Active Membranes: Attacking NP-complete Problems. *Journal of Automata, Languages and Combinatorics* **6**, 75-90 (2001).
12. Gh. Păun, G. Rozenberg, A. Salomaa (Eds.). *The Oxford Handbook of Membrane Computing*, Oxford University Press (2010).
13. M.J. Pérez-Jiménez, A. Riscos-Núñez, A. Romero-Jiménez, D. Woods. Complexity-Membrane Division, Membrane Creation. In [12], 302–336.
14. A.E. Porreca, A. Leporati, G. Mauri, C. Zandron, Sublinear-Space P Systems with Active Membranes. *Lecture Notes in Computer Science* **7762**, 342–357 (2013).

Discrete and Continuous Time High-Order Markov Models for Software Reliability Assessment

Vitaliy Yakovyna and Oksana Nytrebych

Software Department, Lviv Polytechnic National University, Lviv, Ukraine
vitaliy.s.yakovyna@lpnu.ua, ksenija.volynj@gmail.com

Abstract. Due to the critical challenges and complexity of modern software systems developed over the last decade, there has arisen an ever increasing attention to look for products with high reliability at reasonable costs. Software development process moves toward component-based design, and architecture based approach in software reliability modeling is widely used. However, in lots of models for software reliability assessment the assumption of independent software runs is a simplification of real software behaviour. This paper describes two software reliability models that use high-order Markov chains thus taking into account dependencies among software component runs for more accurate software reliability assessment. The efficiency and accuracy of developed models is investigated by the example of several software products. It is shown that using the software reliability models based on the high-order Markov chains results in the software reliability assessment accuracy up to 10–20%.

Keywords. Software reliability, architecture-based model, high-order Markov chains.

Key Terms. SoftwareComponent, SoftwareSystem, MathematicalModel.

1 Introduction

Computer systems are widely used in modern industry for control and automation purposes. All of these systems are controlled by software. Thus, software is used in air traffic control, nuclear power plants, automated patients monitoring etc. Therefore high requirements for software reliability are demanded because failures of such systems can lead not only to significant financial losses, but also threaten human life and health. Although techniques that allow assessing and ensuring the specified hardware reliability requirements have been developed, there are no common approaches for software systems assessment.

According to STD-729, software reliability is defined as the probability of failure-free software operation for a specified period of time in a specified environment [1, 2]. Although Software Reliability is defined as a probabilistic function, and comes with the notion of time, it should be noted that, contrast to traditional hardware relia-

bility, software reliability is not a direct function of time [1]. Electronic and mechanical parts may become "old" and wear out with time and usage, but software will not rust or wear-out during its life cycle. Software will not change over time unless intentionally changed or upgraded.

The history of software reliability assessment methods and tools began in the 60s of last century. A number of researchers [3-7] worked on issues of development and research of software reliability analysis and assessment models and methods that would allow reducing the costs required for software testing stage. New software reliability assessment models that reflect the internal structure of the application and interaction of its components [6], called architectural-based models, are being developed because of increasing complexity of software systems and the result of expanding their functional purpose.

In known models based on architectural approach [6, 7] the theory of first order Markov chains is used with the assumption that the software components runs are independent. This assumption is not always true due to the complexity of modern software architecture and huge set of usage scenarios this assumption [8, 9]. Therefore, for the development of adequate software reliability assessment models, which will improve the testing process (e.g. allow to reduce the needed resources), one should consider the high-order Markov chains, which allow to take into account the interdependence of components runs [10].

This paper describes the software reliability assessment models based on architectural approach, which use discrete and continuous time high-order Markov chains and their comparison based on real world data. Nowadays architecture-based software reliability models have been well studied in theory, but there is lack of papers describing their practical applications. Developing of high-order models for software reliability assessment is not described enough in literature as well. Thus, the development of new architecture-based software reliability assessment models that take into account the dependencies among software component runs along with their application to real world data is still a problem waiting for solution.

2 Software Reliability Model with Discrete Time High-Order Markov Chain

This model considers absorbing Markov process that implies the existence of one or more absorbing states, i.e. states that, once entered, cannot be left.

The developed software reliability assessment model [10] is hierarchical, that is initially software architecture parameters are calculated based on software usage model [11] using the theory of Markov processes, and then the behaviour of each component failures is taken into account.

The discrete time HOMC software reliability model can be described using the following components – $\{C_i\}$ is the graph with vertices corresponding to software components, while edges indicates the program control flow ($i = \overline{1, N}$, where N is the number of program components); $\mathbf{P} = \{p_{ij..kl}\}$ is the high-order transition proba-

bility matrix ($p_{ij..kl}$ – transition probability from component i to component l depending on being in previous K components); $\mathbf{Q} = \{q_{ij..kl}\}$ is the initial probability vector; $\lambda_i(t)$ is the failure rate of i-th software component.

According to this model the reliability of the whole system is calculated as

$$R = \prod_{l=1}^{N} R_l \quad . \tag{1}$$

In turn, the reliabilities of each component (R_l) using high-order Markov chains are calculated by

$$R_l = \exp(-\int_0^{\sum_{ij..k} V_{ij...kl} t_{ij...kl}} \lambda_l(t)dt) \quad . \tag{2}$$

To calculate $V_{ij..kl}$ – the expected number of visits a component l depending on being in previous K components – one has to solve the following system of linear equations:

$$V_{j..kl} = q_{ij..k} + \sum_{i=1}^{N-1} V_{ij..k}\, p_{ij..kl} \quad , \tag{3}$$

here $t_{ij..kl}$ denotes the time spent at the component l depending on being in previous K components.

3 Software Reliability Model with Continuous Time High-Order Markov Chain

Using discrete-time model has a number of significant simplifications and restrictions. Thus, this model takes into account only the number of visits to i-th component without taking into account of the distribution function of this random variable (while taking it into account the expected value should be used). In addition, it is clear that at any given time t the software failure is generally caused by the failure of the software component, which is executed at a given time (in case of sequential connection of software components in Reliability Block Diagram). Thus, to increase the degree of adequacy of the software reliability architectural model the continuous time Markov chains should be used.

The continuous time HOMC software reliability model can be described using the following components – $\{C_i\}$ is the graph with vertices corresponding to software components, while edges indicates the program control flow ($i = \overline{1, N}$, where N is the number of program components); $\mathbf{A} = \{a_{ij}\}$ is the high-order transition probabil-

ity matrix $i, j = \overline{1, N}$ (the values of matrix elements a_{ij} depends on the way of getting into the state i); $\mathbf{P} = \{p_i(t)\}$ is the probability vector, where $p_i(t)$ is the probability being in state C_i at time t; $\lambda_i(t)$ is the failure rate of i-th software component.

In this case, the failure rate of a software system consisting of *N* components can be written as [12]

$$\lambda(t) = \sum_{i=1}^{N} p_i(t) \cdot \lambda_i(t) \quad (4)$$

,

here $\lambda_i(t)$ is the failure rate of i-th component, $p_i(t)$ is the probability of i-th component execution at time t.

Components failure rates $\lambda_i(t)$ can be obtained from the results of unit testing using known software reliability models, for example ones based on an inhomogeneous Poisson process [13].

If the flow control between the components of the software is presented as a Markov process with continuous time, assuming that the i-th state of the process is the execution of i-th component, the time dependences of the probabilities being in i-th state ($p_i(t)$) can be obtained by solving the system of equations of the Kolmogorov–Chapman [14] for this process:

$$\frac{dp_i(t)}{dt} = -\sum_{j \in S} \omega_{ij}(t) p_j(t) + \sum_{j \in S} \omega_{ji}(t) p_j(t), i \in S \quad (5)$$

,

here $\omega_{ij}(t)$ is the transition intensity from component *i* to component *j* at time *t*, and *S* denotes the set of all system states. In general the transition intensities $\omega_{ij}(t)$ depends on the transition probabilities a_{ij} and could be calculated from latter.

To take into account the interdependence of execution of software component (and consequently changes of transition probability from i-th state) depending on the way of getting to the current state, the high-order Markov chain should be used (the order K of the model determines the accounted length of the path). In a case of high-order Markov chains the actual problem is to calculate the transition probabilities depending on the program control flow background.

It is well known that a high order Markov chain can be represented as a first order chain by appropriate redefining the state space [15]. For software implementation of models that use high-order chains, it is necessary to have the formalized algorithm for this representation. Using the analogy with the known Erlang phase method [16] a high-order Markov process can be represented as an equivalent first-order process with additional virtual states. Each state of the original graph $\{C_i\}$ (geometric dia-

gram, showing the possible states of the system and the possible transitions of the system from one state to another) is split into such number of virtual states as many different paths of length K to this state exist. Thus the problem solution essentially is reduced to using of graph theory. Therefore, to calculate the number m_{ij}^K of chains of *K*-th order from state *i* to state *j* the following expression based on the Floyd method can be used:

$$m_{ij}^K = \sum_{j=1}^{S} \left(m_{il}^{(K-1)} + e_{il} \right) m_{lj}^1 \quad (6)$$

,

here e_{ij} are the elements of the identity matrix.

Using (6) one can build an equivalent graph $\{C_i'\}$ which is a representation of the initial graph $\{C_i\}$, taking into account all K-th order paths to i-th state. This allows to avoid the dependencies of the transition intensities $\omega_{ij}(t)$ on the program control flow background. Then the time dependence of the software component execution probability is obtained by solving the system of Kolmogorov–Chapman equations (5) for the equivalent first order Markov process. Using this dependence in (4) together with the component failure rate, the failure rate of the whole software system can be calculated.

For calculation of software reliability measures, based on the obtained relationship (4), one can use the following relations [14]:

- reliability function *P*(*t*) – the probability that no failure will occur within the time interval [*0, t*] – can be calculated as

$$P(t) = \exp(-\int_0^t \lambda(\tau) d\tau) \quad (7)$$

- mean operating time to failure T_l

$$T_1 = \int_0^\infty P(\tau) d\tau \quad (8)$$

4 Determination of the Markov Chain Order for Software Reliability Assessment

In this paper it is proposed to use AIC and BIC criteria [17] since they are not a hypothesis test and they don't use the significance level. Although these criteria give consistent results and don't depend on the estimated model order. They are efficiently

used for weather forecasting and selection of adequate environmental model [18]. But the usage of these criteria in software reliability modeling still remains unexamined.

In general AIC criterion [19] is calculated by the following expression:

$$AIC = 2k - 2\ln(L), \tag{9}$$

here k is the number of independent parameters in the model, and L is the model's maximum likelihood function.

If the value of this software model's maximum likelihood function is substituted in the expression (9), and instead of parameter k one substitute the number of the K-th order model parameters, which contain N components ($k = N^K(N-1)$), then the following expression can be obtained [20]

$$AIC(K) = 2N^K(N-1) - 2\ln(\prod_{i,j,..,k,l} p_{ij..kl}^{n_{ij..kl}}), \tag{10}$$

here $n_{ij..kl}$ is the number of transitions from component i to component l depending on being in previous components ($i \to j \to ... \to k \to l$) in observed sequence and $p_{ij..kl}$ is the transition probability in this sequence.

As it can be seen from (10), this criterion is independent of sample size. Therefore, AIC criterion is used in the case of a large sample size (observable sequence) when the software is tested many times and the sequence of software component runs is logged.

An alternative to AIC criterion is BIC [21], which takes into account the number of elements in observable sequence and expressed as

$$BIC(K) = -2\ln(L) + \ln(n)k. \tag{11}$$

In the case of using the BIC criterion for the Markov chain optimal order determination in the case of software reliability modeling, the expression similar to AIC criterion (10) is obtained:

$$BIC(K) = 2N^K(N-1)\ln(n) - 2\sum_{i,j,..,k,l} n_{ij..kl}\ln p_{ij..kl}\ . \tag{12}$$

It is worth noted that the BIC criterion should be used at the small sample size of empirical data concerning the software usage, because it imposes stronger penalties even when the sample size n>8.

Thus, after determining the order of Markov chain, the well-known classical Markov tools can be used for software reliability assessment as it was described above.

5 Verification of the Models

To study the efficiency of the developed software reliability assessment model that uses the discrete time high-order Markov chains, the reliabilities of the five software systems, developed by one of the authors, were calculated using the first and the high order Markov chain models. The average amount of software components in tested software systems is 10. The results of reliability calculations using the first and the high-order models are shown in Fig. 1 along with the reliability obtained from unit testing data.

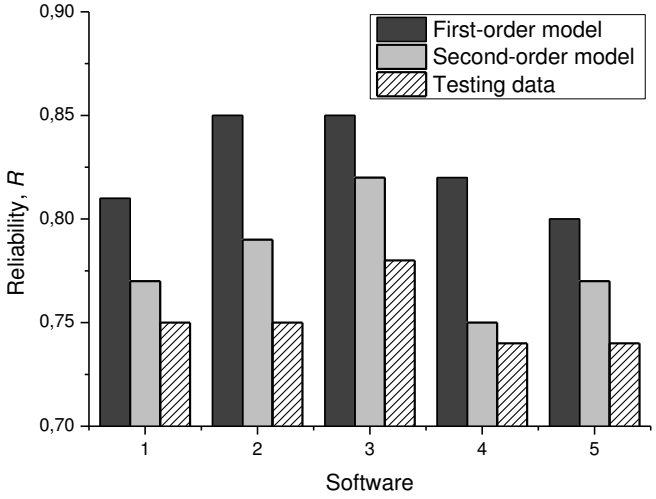


Fig.1. The comparison diagram of the software reliability assessment accuracy using first and high order discrete time models.

As shown in fig. 1, the usage of software reliability assessment model based on discrete time high-order Markov chains makes it possible to increase the software reliability accuracy to 6% even for software systems with a small number of components.

To illustrate the efficiency of continuous time software reliability model, the reliability of the authors' developed program with four components (in this case, each component refers to one of the following program classes – Input, Calculation, Output, Exit) was analyzed. The values of each component failures rates are presented in Table 1.

Table 1. Failure detection frequency for each component

Component	Failures detection frequency
Input	0.11
Calculation	0.18
Output	0.09
Exit	0.01

The probability transition matrix and the initial probability vector of the Markov chain were calculated and are summarized in Table 2 and Table 3 correspondingly.

Table 2. The first order probability transition matrix

Component	*Input*	*Calculation*	*Output*	*Exit*
Input	0.2987	0.66233	0	0.03897
Calculation	0.2666	0.05	0.6333	0.0501
Output	0.5	0	0.3148	0.1852
Exit	0	0	0	1

Table 3. The first order initial probability vector

Components	*Input*	*Calculation*	*Output*	*Exit*
Probability	1	0	0	0

The AIC criterion was used for optimal Markov chain order determination (the sample set contains 188 entries). The AIC values are listed in Table 4.

Table 4. The values of AIC for different chain orders

The process order	**AIC value**
1	223.7
2	203.5
3	352.7
4	1082.8

The initial graph $\{C_i\}$ indicating the program components and control flow, as well as graph $\{C'_i\}$ of the equivalent second order (see Table 4) Markov process are shown in Fig. 2 and 3 correspondingly. The notations in these figures are as follows: **I** correspond to *Input* state, **C** – to *Calculation*, **O** – to *Output*, and **E** to *Exit* state (see Table 2); indexes indicates the previous state in control flow history (thus state $\mathbf{O}_\mathbf{C}$ means that the current state *Output* has been reached from previous state *Calculation*).

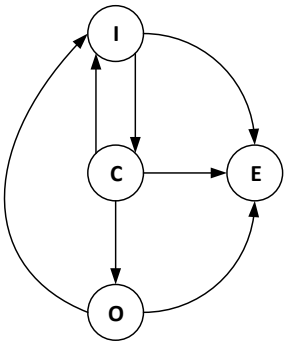


Fig.2. Initial graph, representing the components and control flows of the software used for continuous time reliability model verification.

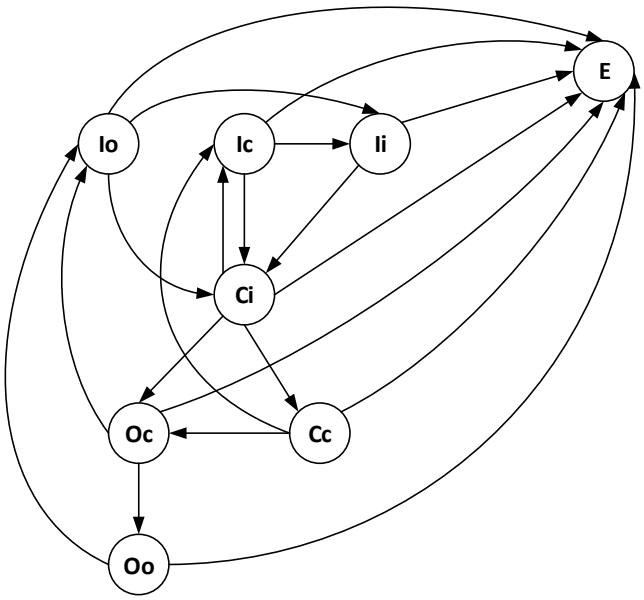


Fig.3. Graph representing an equivalent to the second order Markov chain process for software shown in Fig. 2.

Fig. 4 shows the time dependency of failure rate on tested software system (Fig. 2) obtained from continuous time first- and second-order Markov chains. It's worth noting that for continuous time model simulations time was counted as arbitrary time units (a.u).

As it is seen from this figure the first order model gives slightly (1–5%) reduced failure rate value for small values of time, while it gives significantly (20%) overestimated failure rate value for middle time values range, and when the value of t increases by more than 60 a.u. the difference between values of $\lambda(t)$ obtained from

both models almost tends to be negligible. Reducing the difference of $\lambda(t)$ values, obtained from the first and the second-order models, to zero at high times can be explained by decreasing of the $\lambda(t)$ value itself and by the absence of differences in the software behaviour description by two models in this case. In the case of $t \to \infty$ both models suggest that system is in the "*Exit*" state (see Fig. 2, 3) and, accordingly, its failure rate is limited by this component failure rate. Differences of $\lambda(t)$ behaviour on the initial evolution stages of software system can be entirely explained by differences in the software system behaviour description by different models, where the first-order model ignores the interdependence of software component runs, and therefore the probabilities of components executing are different at given time t for both models. Thus, it could be argued that software reliability model based on the high order Markov chain describes the software system behaviour more adequately and determines its reliability measures more accurately.

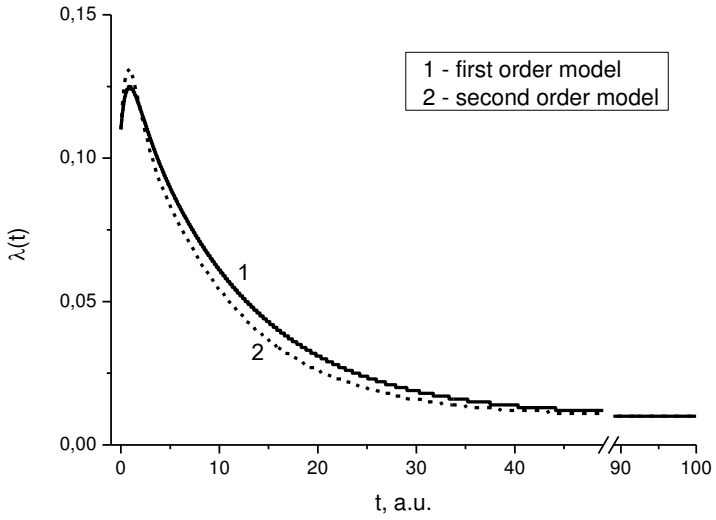


Fig.4. Time dependence of the tested software failure rate, obtained from the first (1) and second-order (2) models.

This conclusion is confirmed by the analysis of software system reliability function time dependence obtained using equation (7) as shown in Fig. 5. Note that the reliability function is calculated using Ukrainian national standards [22] and, as it was indicated in (7), represents the probability of failure free operation during the time interval $(0,t]$, but not exactly at the time t. So, it is evidently that the reliability function is time-decreasing one. As it can be seen from Fig. 5, the difference in the values $P(t)$ increases up to ~20% with time value increasing. This behavior is easily

understood if taking into account the interval estimation of reliability for the range $(0,t]$ [15], and its deviation evidently increases with interval length increasing.

So, we can conclude that ignoring the interdependence of software components runs could results in increasing inaccuracy of software reliability measures estimation up to 20%.

Another measure of reliability, which was used to determine the developed high-order model effectiveness and adequacy, is the mean operating time to failure T_1. The value of mean operating time to failure calculated by the expression (8) is 20.9 time units for first-order model, while the value obtained from the second-order model is 23.3 time units. Obviously the difference between the first and the second-order models is 11.5%, which can be crucial in the reliability analysis of the complex technical systems.

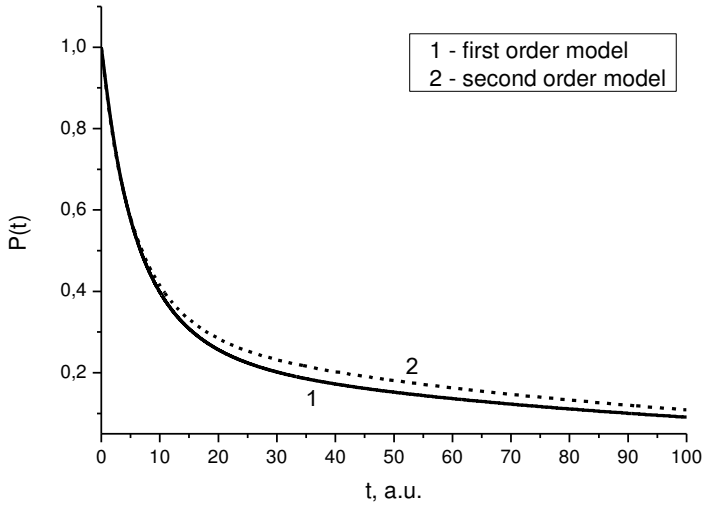


Fig.5. The time dependence of the tested software reliability function, obtained from the first (1) and second-order (2) models.

Therefore, the usage of software reliability models based on the high-order Markov chains even with a small components number (in this case there was 4 component software, see Fig. 2) and complexity (in this case the optimal model order is 2, see Table 4) results in the increase of model adequacy and the software reliability assessment accuracy on 10–20%. The value of such growth of reliability estimation accuracy is especially important in the case of complex hardware and software systems in which the mistake of software component reliability estimation can repeatedly affect the accuracy of whole system reliability assessment. Using the high-order models for more complex software systems could results in more significant improvements of reliability assessment.

6 Conclusions

In this paper the software reliability models using discrete and continuous time high-order Markov chains, which enable consideration the dependencies among software components runs, have been developed. Software reliability assessment based on the high-order discrete time Markov chains allows to increase the software reliability measures accuracy up to 6 %, and high-order continuous time Markov chain – up to 10–20%. The advantage of high-order discrete time Markov chains model for software reliability analysis is small computing resources needed even for software with a lot of components, while the advantages of high-order continuous time Markov chains model are independence from the sampling step (in the case of discrete time model it may cause inaccurate transition probabilities calculation), and also automatic consideration of transitions p_{ii} that avoids unnecessary computations. Practical aspects of estimating input parameters for the models as well as studying the dependence of the software reliability on its characteristics (components number, the average number of variables in the component, components cohesion etc.) will be described elsewhere.

References

1. Lyu, M. R.: Handbook of Software Reliability Engineering. McGraw-Hill, New York, U.S.A. and IEEE Computer Society Press, Los Alamitos, California, U.S.A. (1996)
2. Standard Glossary of Software Engineering Terminology, STD-729-1991 (1991)
3. Musa, J.D.: Validity of Execution Time Theory of Software Reliability. IEEE Trans. on Reliability 3, 199–205 (1979)
4. Goel, A. L., Okumoto K.: Time-Dependent Error Detection Rate Model for Software and other Performance Measures. IEEE Trans. on Reliability R-28, 206–211 (1979)
5. Xie, M.: Software Reliability Modelling. World Scientific, Singapore (1991)
6. Goševa-Popstojanova, K., Trivedi Kishor S.: Architecture-Based Approach to Reliability Assessment of Software Systems. Performance Evaluation, 45, 179–204 (2001)
7. Gokhale, S.S., Wong, W.E., Horgan, J.R., Trivedi Kishor, S.: An Analytical Approach to Architecture-Based Software Performance Reliability Prediction. Performance Evaluation 58(4), 13–22 (2004)
8. Takagi, T., Furukawa, Z., Yamasaki, T.: Accurate Usage Model Construction Using High-Order Markov Chains. In: Supplementary Proc. 17th Int. Symposium on Software Reliability Engineering, pp.1–2 (2006)
9. Burkhart, W., Fatiha, Z. (Eds.): Testing Software and Systems. Proc. 23rd IFIP WG 6.1 Int. Conf. LNCS, vol. 7019 (2011)
10. Yakovyna, V., Serdyuk, P., Nytrebych, O., Fedasyuk, D.: High-Order Markov Chains Usage in Software Reliability Analysis. Bulletin of Lviv Polytechnic National University 771, 209–213 (2013) (in Ukrainian)
11. Fedasyuk, D., Yakovyna, V., Serdyuk, P., Nytrebych O.: Variable State-Based Software Usage Model Based on its Variables. Econtechmod 3, 15–20 (2014)
12. Yakovyna, V., Masyukevych, V.: The Model for Software Reliability Estimation Using High-Order Continuous-Time Markov Chains. In: Proc.9th Int. Scientific and Technical Conference on Computer Science and Information Technologies, pp. 83–86 (2014)

13. Seniv, M., Yakovyna, V., Chabanyuk, Ya., Fedasyuk, D.: The Method of Reliability Prediction and Estimation Based on Model with Dynamic Index of Project Size. Computing 10, 97–107 (2011) (in Ukrainian)
14. Polovko A., Gurov, S.: Fundamentals of the Reliability Theory. Practical work. BHV-Peterburg (2006) (in Russian)
15. Markov Models, Hidden and Otherwise, http://kochanski.org/gpk/teaching/0401Oxford/HMM.pdf
16. Volochiy, B.Yu., Ozirkovskii, L.D., Kulyk, I.V.: Formalization of Discrete-Continuous Stochastic Systems Model Building using Erlang Phases Method. Vidbir i Obrobka Informatsii 36, 39–47 (2012) (in Ukrainian)
17. Burnham, P.: Model Selection and Multimodel Inference: a Practical Information-Theoretic Approach. Springer, Heidelberg (2002)
18. Liu, T.: Application of Markov Chains to Analyze and Predict the Time Series. Modern Applied Science 4(5), 162–166 (2010)
19. Akaike, H.: A New Look at the Statistical Model Identification. IEEE Trans. Auto. Control 19(6), 716–723 (1974)
20. Tong, H.: Determination of the Order of a Markov Chain by Akaike's Information Criterion. J. Appl. Probability 12, 488–497 (1975)
21. Schwarz, G.: Estimating the Dimension of a Model. Annals of Statistics 6(2), 461–464 (1978)
22. Dependability of Technics. Terms and Definitions, State Standard of Ukraine DSTU-2860-94 (1996)

Evolution of Software Quality Models: Green and Reliability Issues

Oleksandr Gordieiev[1], Vyacheslav Kharchenko[2] and Mario Fusani[3]

[1] University of Banking of the National Bank of Ukraine, 1 Andriivska Street, Kyiv, Ukraine
alex.gordeyev@gmail.com
[2] National Aerospace University «KhAI», 17 Chkalova Street, Kharkiv, Ukraine
V.Kharchenko@khai.edu
[3] ISTI-CNR, System and Software Evaluation Center, Via Moruzzi, 1 56124 Pisa, Italy
mario.fusani@isti.cnr.it

Abstract. The group of attributes (characteristics, requirements) related to green software is essential part of software quality model. It consists of the two main attributes as a resources (energy) saving and sustainability. Evolution of software quality models is analyzed in context of green and reliability. In particular, well known software quality models beginning from on the first McCall's model (1977) to models described in standards ISO/IEC9126 (2001) and ISO/IEC25010 (2010) are analyzed according with green and reliability issues. Comparing of the software quality models are carried out using a special metrics of complexity and technique considering the number of levels and attributes and their semantics. Prediction of complexity for the next software quality model (2020) is fulfilled and variants of green software attributes inclusion in model are proposed.

Keywords. software quality model, green software, software reliability, evolution analysis, metrics, ISO/IEC9126, ISO/IEC25010, structure-semantic analysis.

Key Terms. Model, Reliability, Requirement, SoftwareSystem.

1 Introduction

1.1 Motivation and Work Related Analysis

A set of Software Quality Models (SWQM) has been introduced during evolution of software engineering [1]. Software quality is a degree to which a software product satisfies stated and implied needs when used under specified conditions [2]. Software Quality Model (SWQM) is usually defined as a set of characteristics and relationships between them which actually provide the basis for specifying the requirements of quality, evaluating quality and comparing of SWQMs [3-9]. There are a lot of the models suggested during «software engineering era» [10]. Some of SWQM, described in IEEE, ISO, IEC standards, became well-known and can be called basic. New

significant SWQM appear just about once in 10 years. The characteristics and subcharacteristics set and structure (graph-based hierarchy and semantic content) of such SWQMs are changed [11-14]. Generally, these sets are extended and the next SWQM becomes more and more complicated. Changing's of SWQMs are caused by evolution of technologies, new challenges in software engineering and so on.

One of the challenges is development of energy-saving (green) information technologies. It has been caused by appearance of a concept «green software» [15]. Gist of «green software» (GSW) in a broad sense is described by the following words: «decrease» (energy or other resources consumption), «don't do much harm and preserve» (energy, resources, environment) and «improve» (make environment more comfortable and safe). More wide aspects and directions of green and safe/reliable computing are discussed in [16,17].

«Green» characteristics for software are resources saving and sustainability, which were not explicitly defined in well known SWQMs described by standards ISO/IEC9126 [18], ISO/IEC25010 [2]. Analysis of [3,4,6-8] allowing to conclude that SWQMs do not include such characteristics in explicit form.

Taking into consideration the prerequisites for emergence of green characteristics in future SWQMs in direct form we analyze the evolution of the characteristics associated with GSW for existing quality models and try to predict their changing. The analysis will allow defining tendencies of green characteristics and suggesting variants of including some in future SWQMs.

1.2 Goal and Approach

A **goal of the paper** is carrying out of analysis of known software quality models and their development in context of GSW and software reliability. We aim to investigating SWQMs using metric-based approach to assess "weights" of different software quality attributes, first of all, green and reliability characteristics, changing of the weights during evolution of the models and to predict their changing in future.

Stages of the research are the following:

1. Determination of occurrence rates for different SWQM attributes (characteristics at the first level of hierarchy and subcharacteristics at the second one) in different quality models;
2. Selection and analysis of SWQM characteristics which are implicitly associated with green software;
3. Analysis of SWQMs in context green software and reliability by use of complexity metrics and calculation of corresponding weights for attributes;
4. Research of relationship/dependency between metric values for green software, reliability and the years of emergence for known basic SWQMs;
5. Calculation of complexity metric for using results of SWQMs relationship/dependency comparison, described in [11];
6. Calculation of complexity metric for green and reliability attributes of new SWQMs using function describing of dependency between metric values and years of SWQMs emergence;
7. Analysis of SWQM in use in context of green software and definition of possible variants of inclusion of green attributes in new models.

2 SWQM Analysis in Context of Green Software and Reliability

2.1 Analyzed Models

Let's select and analyse SWQM characteristics which can be implicitly associated with green software and reliability. The results of analysis are shown in Table 1 and Table 2 for green characteristics and reliability characteristics correspondingly. Numeration of the characteristics corresponds with their "places" in hierarchy of SWQMs.

Table 1. SWQM characteristics associated with GSW.

№	SWQMs (years)	GSW characteristics
1.	McCall (1977)	4. Efficiency
		4.1 Execution efficiency
		4.2 Storage efficiency
2.	Boehm (1978)	2.2 Efficiency
		2.2.1 Accountability
		2.2.2 Accessibility
3.	Carlo Ghezzi (1991)	–
4.	FURPS (1992)	4 Performance
		4.1 Velocity
		4.2 Efficiency
		4.3 Availability
		4.4 Time of answer
		4.5 Time of recovery
		4.6 Utilization of resources
		1.2 Capacity
5.	IEEE (1993)	1 Efficiency
		1.1 Temporal efficiency
		1.2 Resource efficiency
6.	Dromey (1995)	2.2 Efficiency
7.	ISO 9126-1 (2001)	4 Efficiency
		4.1 Time behavior
		4.2 Resource utilization
8.	QMOOD (2002)	6 Effectiveness
9.	ISO 25010 (2010)	2 Performance efficiency
		2.1 Time behavior
		2.2 Resource utilization
		2.3 Capacity

Table 2. Reliability characteristics of SWQM.

№	SWQMs (years)	Reliability characteristics
1.	McCall (1977)	2. Reliability
		2.1 Accuracy
		2.2 Error tolerance
		2.3 Consistency
2.	Boehm (1978)	2.1 Reliability
		2.2.1 Self contentedness
		2.2.2 Integrity
		2.2.3 Accuracy
3.	CarloGhezzi (1991)	3. Reliability
4.	FURPS (1992)	3. Reliability
		3.1 Frequency and servity of failures
		3.2 Recoverability
		3.3 Time among failures
5.	IEEE (1993)	2. Reliability
		2.1 Non deficiency
		2.1 Error tolerance
		1.3 Availability
6.	Dromey (1995)	1.2 Reliability
7.	ISO 9126-1 (2001)	2. Reliability
		2.1 Maturity
		2.2 Fault tolerance
		2.3 Recoverability
8.	QMOOD (2002)	-
9.	ISO 25010 (2010)	5. Reliability
		5.1 Maturity
		5.2 Availability
		5.3 Fault tolerance
		5.4 Recoverability

To assess "weights" of green characteristics the technique of SWQM structure-semantic analysis (SSA-technique) can be applied [11]. The technique describes quality models as a facet-hierarchy structure (graph). Nodes corresponds quality attributes and links take into account hierarchy dependencies. To briefly characterize the proposed analysis technique, let us introduce some initial terms:

- conceptual model is a model which a model under study is compared with;
- model under study is a model which is compared with a conceptual model;
- characteristic under study is a conceptual model characteristic which is compared with model under study characteristics.

2.2 Metrics

SSA-technique is based on comparing a model under study with the conceptual model, i.e. every SW Quality Model is compared with the conceptual model. So, the analysis is equivalent to semantic comparing characteristics and subcharacteristics of a model under study and the conceptual model with regard to their structures. Selecting a reference model is usually performed by an expert who has relevant experience and qualifications.

At the following stage comparison of models among themselves should be performed. The simplest and most obvious metrics are offered. Hierarchy of these metrics is presented in Fig. 1. The metrics are used to compare models with reference model bottom up, i.e. first at the level of subcharacteristics (subcharacteristics matching metric SMM, cumulative subcharacteristics comparison metric CSCM, characteristics matching metric CMM), then at the level of characteristics (cumulative matching characteristics metric CMCM) and finally at the level of models as a whole (cumulative software quality models comparison metric CSQMCM).

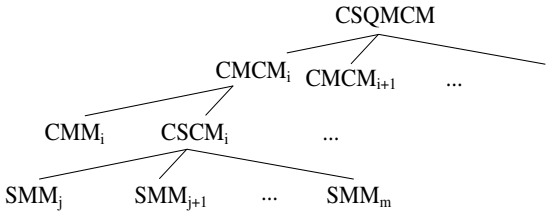


Fig. 1. Metrics hierarchy.

Features of the metrics are the following:

– subcharacteristic matching metric (SMMj). Every subcharacteristic match value is identified as SMMj = 0,5 / number of reference (conceptual) model elements subcharacteristics of the characteristic under study. Weights of characteristics are not considered when calculating metrics;

– cumulative subcharacteristics comparison metric (CSCM) is evaluated as a sum of SMM:

$$CSCM_i = \sum_{j=1}^{k} SMM_j ; \quad (1)$$

– characteristics matching metric (CMM) takes the value of 0.5 in case of matching or 0 if the characteristics are different;

– cumulative matching characteristics metric (CMCM) is calculated as a sum of CMM metric and $\sum_{j=1}^{k} CSCM_j$:

$$CMCM_i = CMM_i + \sum_{j=1}^{k} CSCM_j ; \quad (2)$$

– cumulative software quality models comparison metric (CSQMCM) is calculated according to the formula:

$$CSQMCM_i = \sum_{j=1}^{n} CMCM_j \quad (3)$$

2.3 Results of SWQM Analysis in Context of Green Software and Reliability Characteristics

Let us conduct SW QM analysis and first of all, define the reference (conceptual) model. SW Quality Model ISO/IEC 25010 will be considered as uppermost and etalon regarding to all other models. It is the newest introduced model and takes into account main modern software peculiarities in point of view quality evaluation. This model is described by international standard of top level.

According with results of analysis CMCM is calculated for set of characteristics presented in Table 1. The results of calculation are shown in Table 3 (Chs – characteristics, SChs – subcharacteristics) for GSW characteristics and Table 4 for reliability characteristics.

The histogram of CMCM values for software quality models is presented on Fig. 2. An abscissa axis corresponds to years of SWQM emergence. Initial point (year) is 1970 (as a first year after 1968 which is multiple of a ten years).

CMCM values will be further represented and analysed only for so-called basic SWQMs [18]. Basic models were selected considering their support by standards, the international reputation and application. The models of McCall and Boehm are similar, hence first one was selected. Hence, the models of Boehm, Ghezzi, FURPS, Dromey, QMOOD were excluded (Fig. 3).

The analytical dependency between SWQM appearance year (X axis) and CMCM value (Y axis) for characteristics associated with GSW may be represented by regressive liner function:

$$y = ax+b, \quad (4)$$

where x – variable, a and b - regression coefficients. For 1970 year variable (x) has value 0, for 1980 year x=10, for 1990 year x=20, for 2000 year x=30 and for 2010 year x=40.

Linear subjection was chosen by graphic data analysis (Fig. 3). Satisfiability of applying linear subjection is confirmed by coefficient of determination (R2) which equals 0,94.

Table 3. Results of GSW characteristics comparison and CMCM calculation.

Conceptual model (ISO 25010)		**McCall model**				**Boehm model**				**Ghezzi model**			
Chs	SChs	Chs	SChs	CMM	SMM	Chs	SChs	CMM	SMM	Chs	SChs	CMM	SMM
2		4		0,5	0	-	2.2	0	0,5	-	-	0	0
	2.1	-	-	0	0	-	-	0	0	-	-	0	0
	2.2	-	-	0	0	-	-	0	0	-	-	0	0
	2.3	-	-	0	0	-	-	0	0	-	-	0	0
				CMCM=0,5				CMCM=0,5				CMCM=0	
Conceptual model (ISO 25010)		**FURPS Model**				**IEEE Model**				**Dromey model**			
Chs	SChs	Chs	SChs	CMM	SMM	Chs	SChs	CMM	SMM	Chs	SChs	CMM	SMM
2		-	4.2	0	0,5	1	-	0,5	0	-	2.2	0	0.5
	2.1	-	-	0	0	-	-	0	0	-	-	0	0
	2.2	-	4.6	0	0,17	-	1.2	0	0,17	-	-	0	0
	2.3	-	1.2	0	0,17	-	-	0	0	-	-	0	0
				CMCM =0,84				CMCM =0,67				CMCM =0,5	
Conceptual model (ISO 25010)		**ISO 9126 model**				**QMOOD model**							
Chs	SChs	Chs	SChs	CMM	SMM	Chs		SChs		CMM		SMM	
2		4	-	0,5	0	2		-		0,5		0	
	2.1	-	4.1	0	0,17	-		-		0		0	
	2.2	-	4.2	0	0,17	-		-				0	
	2.3	-	-	0	0	-		-		0		0	
				CMCM=0,84						CMCM=0,5			

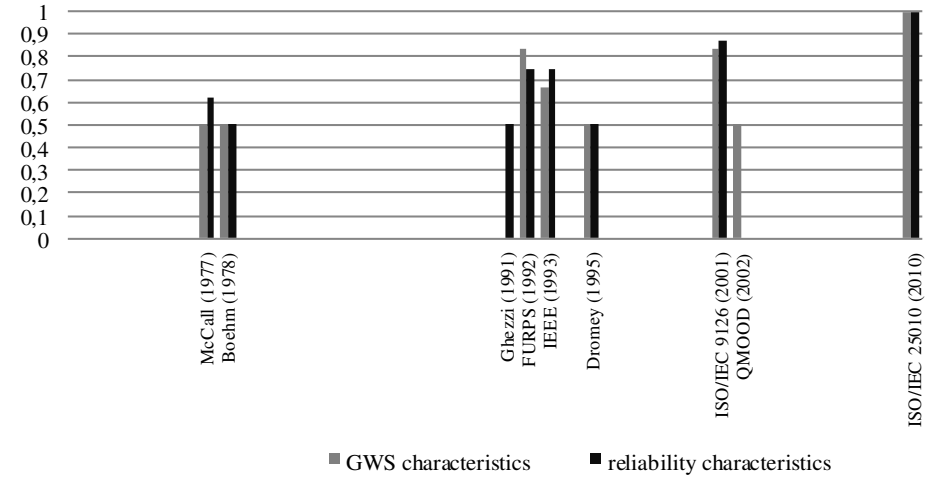


Fig. 2. CMCM values for GSW and reliability characteristics of SWQMs.

Table 4. Results of reliability characteristics comparison and CMCM calculation.

Conceptual model (ISO 25010)		**McCall model**				**Boehm model**				**Ghezzi model**			
Chs	SChs	Chs	SChs	CMM	SMM	Chs	SChs	CMM	SMM	Chs	SChs	CMM	SMM
5		2.	-	0,5	0	-	2.1	0	0,5	3	-	0,5	0
	5.1	-	-	0	0	-	-	0	0	-	-	0	0
	5.2	-	-	0	0	-	-	0	0	-	-	0	0
	5.3	-	2.2	0	0,125	-	-	0	0	-	-	0	0
	5.4	-	-	0	0	-	-	0	0	-	-	0	0
				CMCM=0,625				CMCM=0,5				CMCM=0,5	
Conceptual model (ISO 25010)		**FURPS Model**				**IEEE Model**				**Dromey model**			
Chs	SChs	Chs	SChs	CMM	SMM	Chs	SChs	CMM	SMM	Chs	SChs	CMM	SMM
5			3.	-	0,5	2		0,5	0		1.2,2.3, 3.4,4.4	0	0,5
	5.1	5.1	-	-	0	-	-	0	0	-	-	0	0
	5.2	5.2	-	4.3	0	-	2.3	0	0,125	-	-	0	0
	5.3	5.3	-	-	0	-	2.2	0	0,125	-	-	0	0
	5.4	5.4	-	3.2	0	-	-	0	0	-	-	0	0
				CMCM =0,75				CMCM =0,75				CMCM =0,5	
Conceptual model (ISO 25010)		**ISO 9126 model**				**QMOOD model**			
Chs	SChs	Chs	SChs	CMM	SMM	Chs	SChs	CMM	SMM
5		2	-	0,5	0	-	-	0	0
	5.1	-	2.1	0	0,125	-	-	0	0
	5.2	-	-	0	0	-	-	0	0
	5.3	-	2.2	0	0,125	-	-	0	0
	5.4	-	2.3	0	0,125	-	-	0	0
				CMCM=0,87				CMCM=0	

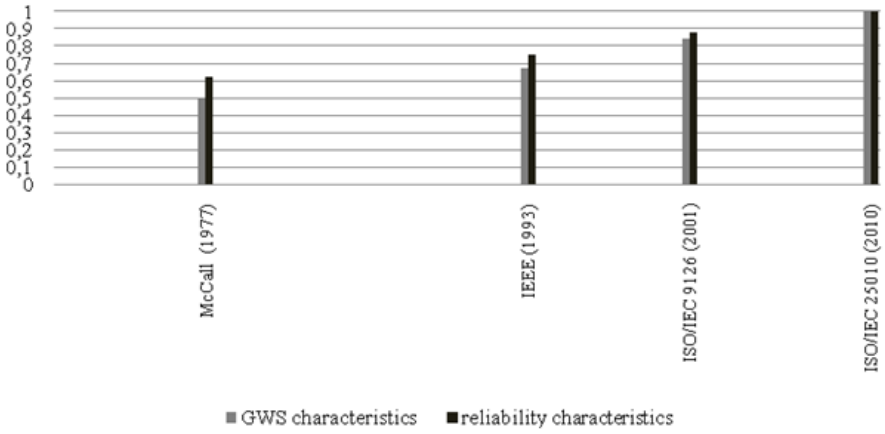


Fig. 3. CMCM values for GSW and reliability characteristics of basic SWQMs.

The values of parameters a and b can be calculated using Least Square Method:

$$a = \frac{n\sum_{i=0}^{n} x_i y_i \quad \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x^2 \quad (\sum_{i=1}^{n} x_i)^2}, \tag{5}$$

$$b = \frac{\sum_{i=1}^{n} y_i \quad a\sum_{i=1}^{n} x_i}{n}. \tag{6}$$

As a result a = 0.0146, b = 0.4108 and function:

$$y = 0.0146x+0.4108. \tag{7}$$

The obtained function may be called a law of increasing of characteristics associated with GSW for SWQM.

The similar dependency can be obtained for reliability characteristics. In this case a = 0.011, b = 0.5 and function:

$$y = 0.011x+0.5. \tag{8}$$

Formulas 7 and 8 illustrate a tendency of SWQMs characteristics/ subcharacteristics changes. Analysis of dependencies (Fig.3) allows concluding that weights of green and reliability characteristics became equal in 2010 (the standard ISO/IEC 25010). Hence, since first SWQMs the characteristics/ subcharacteristics related to green attributes have faster dynamics of increasing.

3 Development of SWQM in Context of Green Software

We can assume that the next general SWQM will include GSW characteristics in an explicit form. Let's analyse SWQM evolution tendency in context GSW as a whole. CSQMCM for SWQM may be calculated as shown in formula (3). It may be appeared for future model (2020 year). In compliance with [11] and basing on the analytical relationship between SWQM appearance year (X axis) and CSQMCM value (Y axis) the following formula may be obtained:

$$y=0,153x+1,363. \tag{9}$$

Besides, considering that each new SWQM approved as a standard is received about once per 10 years, and that the last model was introduced by the standard ISO/IEC 25010 appeared in 2010 the prediction of the CSQMCM value can be done. With this in mind:

$$CSQMCM = 0,153*50 + 1,363= 9,013. \tag{10}$$

CSQMCM values change is illustrated in Fig. 4 as a histogram for the well known base SWQM as columns of gray and subsequent SWQM 2020 as a column of light gray column.

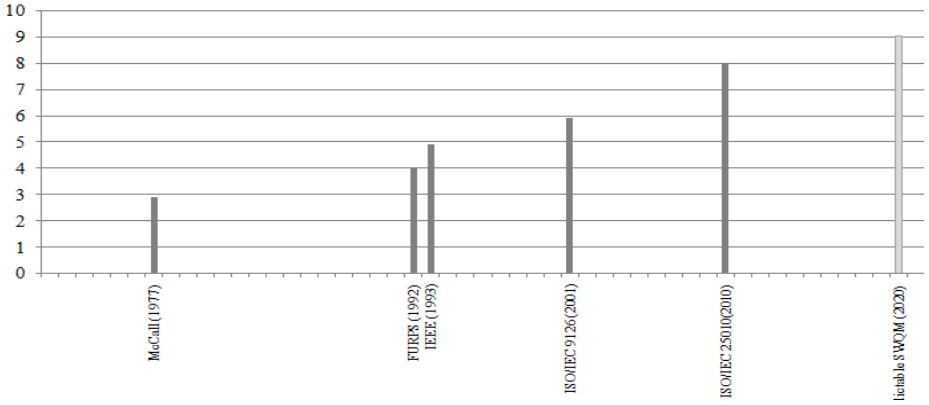


Fig. 4. CSQMCM values for known and predictable SWQMs.

According to the obtained dependence (4) CMCM for green software characteristics is calculated for predictable SWQM 2020 (Fig. 5).

$$y=0,0146*50+0,4108=1,1408. \quad (11)$$

And CMCM for reliability characteristics is calculated for predictable SWQM 2020 (Fig. 5).

$$y = 0,011*50+0,5=1,05. \quad (12)$$

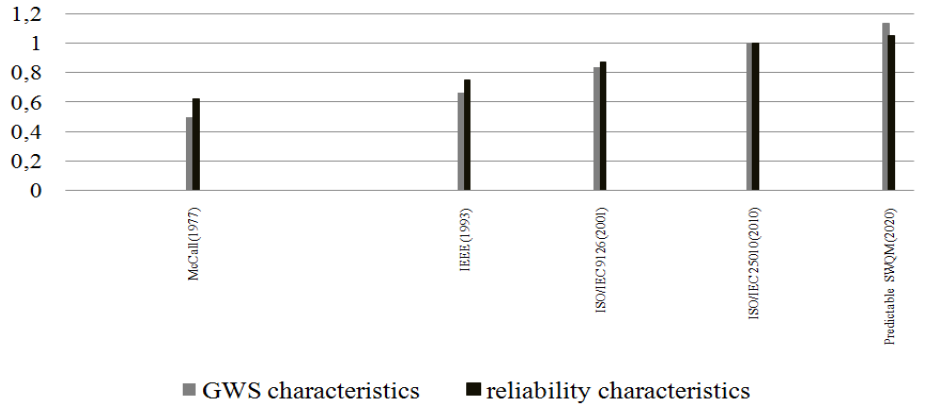


Fig. 5. CMCM values for reliability characteristics and green characteristics for basic SWQMs.

CMCM values of SWQM 2020 for characteristics associated with «green» software exceed the value of the same metric for SWQM ISO/IEC 25010 by 0.1408.

CMCM values of SWQM 2020 for reliability characteristics exceed the value of the same metric for SWQM ISO/IEC 25010 by 0.05.

Analysis of dependencies (Fig.5) allows predicting that green characteristics number will increase faster comparing with other more conservative characteristics.

4 GSW Oriented ON Extending of SWQMs

Taking into account predictable changing of SWQMs let's analyse how content of such models may be added including software quality models in use.

4.1 Variants of GSW Characteristics Inclusion in SWQM

In the following, possible variants are shown of inclusion of GSW characteristics and its components in a SWQM.

1. GSW characteristic can be introduced in SWQM as a separated characteristic with subcharacteristics *resources saving* and *sustainability*. It should be noted that usually *resources saving* excludes *resource utilization* from *performance efficiency* characteristic (Fig. 6).

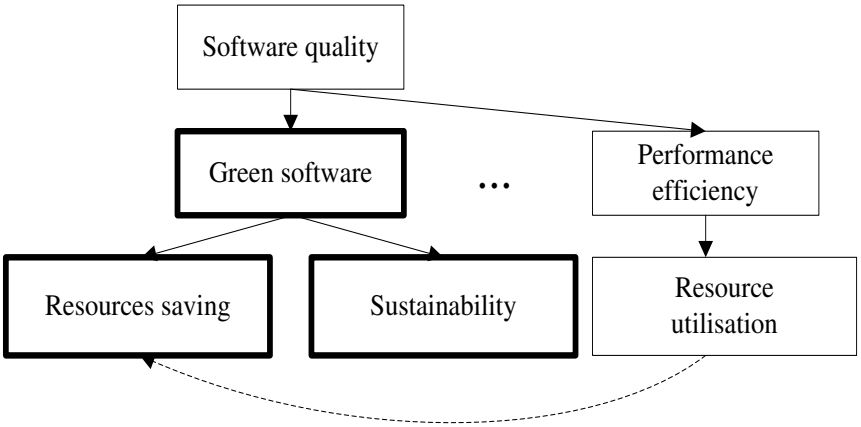


Fig. 6. Green software characteristics in SWQM at the level of characteristics (1).

2. Green software characteristics are not included in SWQM explicitly, but subcharacteristics can go in to SWQM (Fig. 7). *Resources saving* goes in to SWQM as the subcharacteristic in place of *resource utilization*. Subcharacteristic *sustainability* goes in to SWQM as separated characteristic.

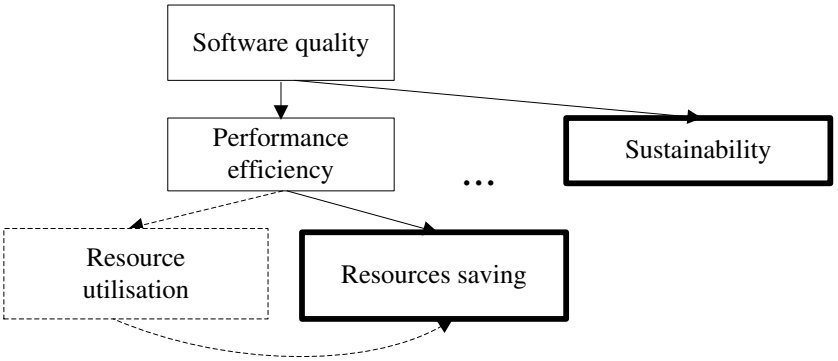


Fig. 7. «Green» software characteristics in SWQM at the level of characteristics and subcharacteristics (2).

3. GSW characteristic cannot be explicitly included in SWQM, but subcharacteristics can be explicitly included (Fig. 8). *Resources saving* is included in SWQM as subcharacteristic in place of *resource utilization. Sustainability* is included in SWQM as subcharacteristic to characteristic *security*.

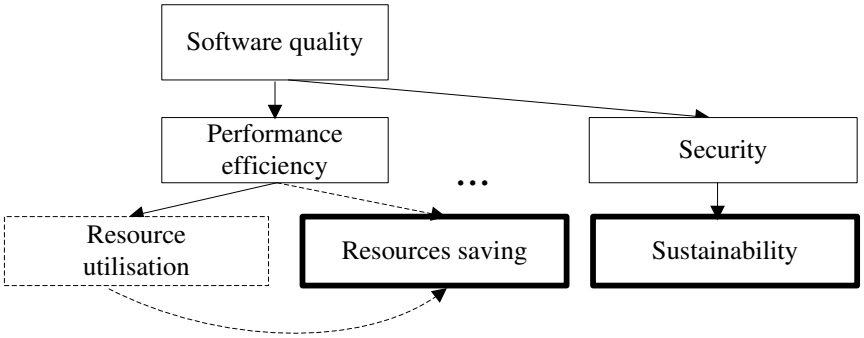


Fig. 8. Green software characteristics in SWQM at the level of subcharacteristics (3).

4.2 SWQM in Use. Analysis in Context of GSW

The standards ISO/IEC9126 and 25010 describe a separate type of models - software quality models in use (SWQM-U). SWQM-U is a capability of the software product to enable specified users to achieve specified goals with effectiveness, productivity, safety and satisfaction in specified contexts of use [18]. The SWQM-Us include characteristics, which can be associated with GSW subcharacteristics, in particular resources saving and sustainability:

– **for SWQM-U, ISO/IEC 9126**: *resources saving – productivity; sustainability – safety. Productivity* is a capability of the software product to enable users to expend appropriate amounts of resources in relation to the effectiveness achieved in a specified context of use. *Safety* is a capability of the software product to achieve acceptable levels of risk of harm to people, business, software, property or the

environment in a specified context of use. Risks are usually a result of deficiencies in the functionality (including security), reliability, usability or maintainability;

– **for SWQM-U, ISO/IEC 25010**: *resources saving – efficiency*; *sustainability – freedom from risk, which include 3* subcharacteristics – *economic risk mitigation*, *health and safety risk mitigation and environmental risk mitigation. Efficiency* is a ratio of expended resources to the accuracy and completeness with which users achieve goals. *Freedom from risk* is a degree to which a product or system mitigates the potential risk to economic status, human life, health, or the environment.

Correlation of SWQM-U characteristics for standards ISO/IEC 9126 and 25010, which are implicitly associated with «green software» and among themselves is shown in Fig. 9.

Thus, GSW related characteristics should be taken into account on development of the next SWQM (SWQM-U) as well.

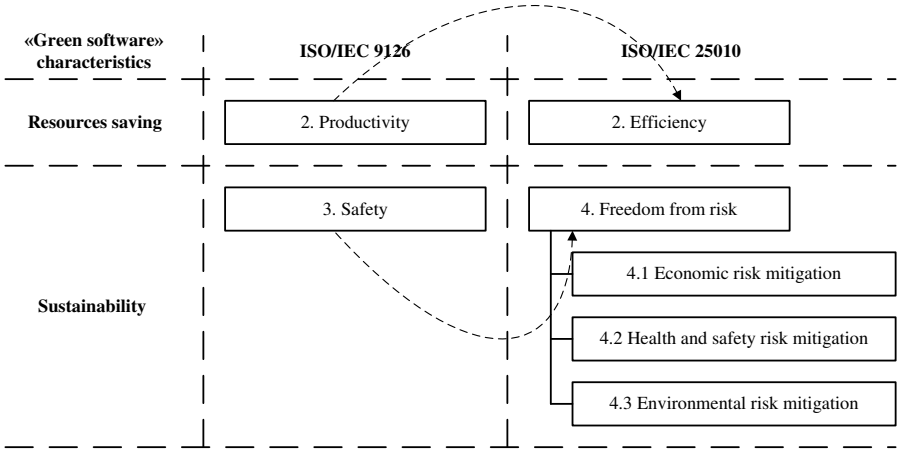


Fig. 9. Correlation of characteristics of SWQM-Us (ISO/IEC 9126 and ISO/IEC 25010) with GSW characteristics.

5 Conclusions

In compliance with SWQM structural and semantic analysis technique we have analyzed SWQM of standards ISO/IEC 9126 and 25010 in context characteristics associated with green software. Using SSA-technique, a relationship between the year of the SWQM appearance and the value of CMCM was obtained and analyzed. Besides, we have calculated the CMCM values for the green software characteristics of the next SWQM, the output of which may be expected in 2020.

It was also obtained the value of metric - CSQMCM for SWQM of 2020, which exceeds the value of this indicator for SWQM ISO/IEC 25010 (Fig. 4). It may be explained by possible inclusion of green software characteristics in SWQM explicitly.

According with results of analysis we can conclude that:

- since first SWQMs the characteristics/ subcharacteristics related to green attributes have faster dynamics of increasing;

- weights of green and reliability characteristics became equal in the standard ISO/IEC 25010;
- it is predicted faster increasing of number green characteristics comparing with other more conservative characteristics.

However, implementation of green characteristics in future quality models should be harmonized with basic attributes such as reliability.

In the future we plan to investigate every SWQM characteristic separately. The data obtained in this case will provide development of a prototype of the new SWQM.

References

1. NATO SCIENCE COMMITTEE Report. Software engineering. Report on a conference sponsored by the NATO SCIENCE COMMITTEE, 136 p., Germany, Garmisch. (1968)
2. International Standard ISO/IEC 25010. Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – System and software quality models, ISO/IEC JTC1/SC7/WG6 (2011)
3. Sanjay Kumar Dubey, Soumi Ghosh, Ajay Rana: Comparison of Software Quality Models: An Analytical Approach. International Journal of Emerging Technology and Advanced Engineering 2(2), 111--119 (2012)
4. Guy Schiavone: A Life Cycle Software Quality Model Using Bayesian Belief Networks. University of Central Florida, Orlando (2006)
5. Jacobson, I., Booch, G., Rumbaugh, J.: The Unified Software Development Process, Addison Wesley Longman, Inc. (1999)
6. Rüdiger Lincke, Tobias Gutzmann, Welf Löwe: Software Quality Prediction Models Compared. International Conference on Quality Software, pp. 82--91 (2010)
7. Stavrinoudis, Xenos: Comparing internal and external software quality measurements. Proccedings of the 8th Joint Conference on Knowledge-Based Software Engineering, pp. 115--124, IOS Press (2008)
8. Amit Sharma, Sanjay Kumar Dubey: Comparison of Software Quality Metrics for Object-Oriented System. Special Issue of International Journal of Computer Science & Management Studies 12, 12--24 (2012)
9. Stefan, Wagner. Software product quality control. Springer (2013)
10. Lami G., Fabbrini F., and Fusani M.: Software Sustainability from a Process-Centric Perspective. Proceedings of EuroSPI 2012, Communication in Computer and Information Science CCIS n. 301, pp. 97--108, Springer, Vienna (Austria) (2012)
11. Gordieiev O., Kharchenko V., Fominykh N., Sklyar V.: Evolution of software quality models in context of the standard ISO 25010. Proceedings of 9th International Conference on Dependability and Complex Systems - DepCoS-RELCOMEX 2014, 30 Jun-4 July, Advances in Intelligent and Soft Computing, pp. 223--233, Springer, Brunow, Poland (2014)
12. Filip Radulovic: A software quality model for the evaluation of semantic technologies. Master thesis. Universidad Politecnica de Madrid Facultad de Informatica, (2011)
13. Rafa E: Al-Qutaish Quality Models in Software Engineering Literature: An Analytical and Comparative Study. Journal of American Science 6(3), 166--175 (2010)
14. Namita Malhotra, Shefali Pruthi: An Efficient Software Quality Models for Safety and Resilience. International Journal of Recent Technology and Engineering (IJRTE) 1(3). 66--70 (2012)
15. San Murugesan, Gangadharan G.R.: Harnessing Green IT. Principles and Practices, UK: Wiley and Sons Ltd. (2012)

16. Kharchenko V., Sklyar V., Gorbenko A., Phillips C.: Green Computing and Communications in Critical Application Domains: Challenges and Solutions. Proceedings of 9th International Conference on Digital Technologies, May, 29-31, 2013, Žilina, Slovakia, pp. 24--29 (2013)
17. Kharchenko V. (editor): Green IT-Engineering. In 2 volumes. Vol.1. Principles, Components and Models - 593 p.; Vol. 2. Systems, Industry, Society. - 628 p. Ukraine: National Aerospace University KhAI (2014)
18. International Standard ISO/IEC9126-1. Software engineering – Product quality – Part 1: Quality, 32 p. (2001)

Service and Business Models with Implementation Analysis of Distributed Cloud Solution

Olga Yanovskaya[1], Maria Anna Devetzoglou[2], Vyacheslav Kharchenko[1,3] and Max Yanovsky[1],

[1]National Aerospace University named after N.E. Zhukovsky "KhAI", Kharkiv, Ukraine
[2]International Creativity Engineering Group, Athens, Greece
[3]Centre for Safety Infrastructure-Oriented Research and Analysis, Kharkiv, Ukraine
[1]{O.Yanovskaya, M.Yanovsky}@csn.khai.edu
[2]M.Devetzoglou@interceg.com
[1,3]V.Kharchenko@khai.edu

Abstract. The service and business models of a Distributed Cloud solution based on peer-to-peer technology are presented in this article, in order to implement the proposed solution. Methods for organizing the interaction between the participants' nodes and nodes that are non-participants in the Distributed Cloud are proposed. Passive replication is used to improve service reliability. A competitive analysis of existing solutions within the scope of a decentralization approach for content sharing is conducted. Average response time to a request for a centralized client-server and distributed Cloud architecture is estimated.

Keywords. Distributed Cloud, Peer-to-Peer Network, Data Center, Participatory Business Model, Service Reliability.

Key Terms. DataCloud, Reliability, Model, Infrastructure, Market.

1 Introduction

The IT industry constantly grows, raising the expectations of big organizations and individuals alike, changing the way things are done nowadays; in fact, without it, it would be impossible to conduct business and human interaction as we know it today would be greatly different across numerous professional and social sectors. At the same time, gradually more people and enterprises access the Internet, adding to the increasing needs of computing and generating data through various devices, at an accelerating rate. This data is required to be stored and handled in order to be secure and available to be retrieved once the user requests it.

To accommodate demand, large, expensive, energy hungry data centers have been built that only powerful company can afford to have. Additionally, they require high costs of maintenance, personnel, power back up systems and space. Located at a physical place, they are vulnerable to local conditions, be it weather phenomena, re-

gional power cuts, earthquakes etc. Furthermore, 2% of global CO_2 emissions are attributed to the ICT industry, a significant part of which is caused by data centers. High investment costs for data centers prevent smaller companies from entering the market, making it necessary to improve service reliability, energy and cost efficiency of Cloud computing infrastructure.

Currently, the concept of P2P technology is not new but its application to cloud computing is at its early stages. However, it is gradually growing as more new companies begin to join the race to provide smarter and cheaper solutions. More specifically, the majority of the companies that are active in P2P cloud technology are strongly focusing on storage and sharing of documents, in order to facilitate storing options and to assist teams or groups to virtually interact through their documents. Additionally, they enhance user experience due to optimized infrastructure, collaboration and sync. BitTorrent goes a step further, by offering information sharing from device to device, skipping the use of cloud [1]. Within this suggestion, we propose a method for Cloud data center architecture modernization [2]. The method assumes implementation of distributional technologies such as peer-to-peer (P2P) networks to Cloud architecture. Distributed Cloud computing, being part of cloud computing, supports customers' needs and provides main cloud benefits. However, it differentiates itself from the existing Cloud Computing in a unique way: it is anthropocentric, revolving around people and their activities, making them sources and consumers at the same time. In addition to its human-centered function, P2P Cloud Computing is a new application of previous technologies, one that can provide both, value and results.

2 State of the Art

According to a Cisco study [3], it is estimated that data traffic will reach 7.7 ZB by 2017 and the need for more websites constantly and steadily rises. This need does not only apply to professionals and big companies that either have the money to allocate the task to a skilled person or have the skills themselves. It also corresponds to a wide number of individuals and companies that, although experienced in a number of sectors, may not have the skills or financial resources to create their content.

Currently, customers deal with scalability issues by using the services of Cloud providers. Studying the pricing of these services [4], it is found that customers may very well enjoy the benefits of the Cloud but at a rather significant cost. Many Cloud providers have joined the sector and provide solutions to customers. For companies that aim to scale up and grow, this can have a heavy toll on their budget and, in more than one cases, budget limitations may actually delay or even prohibit growth. According to the IDC, the cloud software market is forecast to surpass $75 B by 2017, while at the same time, the percentage of IT budget expected to be spent on cloud-based applications and platforms by current organizations within the next 2 years reaches 53.7%. Regarding storage as a service model, several solutions within the use of the decentralization approach were developed for content sharing. The following table depicts competitive products, their key features, their advantages and disadvantages.

Table 1. Competitive analysis of existing solutions within the use of decentralization approach for content sharing

Name	Key features	Advantages	Disadvantages
SpaceMonkey	Storage solution based on P2P technology that saves users' data both locally and remotely in a separate, remote 1 TB hard drive with 2 TB space using as P2P sync [5].	A cross-platform application. Free for the first year.	Expensive device ($795). Annual payment $49. Can only be used as Cloud storage and doesn't support other types of Cloud services.
Project Maelstrom	Web browser based on the Chromium Project that allows access to static web pages using torrent protocol [6].	No limits to file size and transfer speeds.	Can only be applied for static web pages without server part (database, cgi, etc.).
Wuala	Secure Cloud storage, a haven in the Cloud to store customers' files [7].	Components are secure against cryptanalysis (AES-256 for encryption, RSA 2048, SHA-256). System has redundant storage in different locations.	No free subscription. Can only be used as Cloud storage and doesn't support other types of Cloud services.
Sherly	Sharing large files (>20Gb) with secure access control [8].	Robust reporting. Simple access to management controls. Build-in auditing tools. Does not distribute copies of users' data but grants access to it instead. Can be used in both ways with either hardware (Sherlybox) or software.	Can only be used as Cloud storage and does not support other types of Cloud services.
Symform	Customers get 1GB free for every 2GB contributed [9].	A cross-platform application. Free and paid subscriptions.	Can only be used as Cloud storage and does not support other types of Cloud services.
BOINC	Uses the idle time on users' computer (Windows, Mac, Linux, or Android) to perform scientific computing [10].	Free.	Software is for volunteers within the academic society. Does not provide profits to users.

As seen by the table above, almost all of the presented solutions provide users with only one type of Cloud services – storage as a service, meaning they are unsuitable for deploying applications and service delivery. Moreover, the expenses issue remains unsolved.

The aim of this paper is to present the concept of a decentralized cloud architecture based on P2P technology, estimate its availability and to highlight the changes it may bring to business models within the sector. The paper is structured in the following way: section 2 presents the current status of the technology and benefits of Cloud usage. Section 3 describes the proposed solution. In section 4, the service model of the distributed cloud is presented. Sections 5 and 6 examine the response time to a request and the service availability respectively, while section 7 contemplates the business model that may be formed around the proposed idea. Finally, section 8 provides a case study for the implementation of the solution while the conclusion is presented in section 9.

3 Description of the proposed solution

When a company or individual wants to create a website about their activities or themselves, they usually have two options to select from, in order to successfully complete the task.

1. Hire professionals who will handle all the necessary steps, from ensuring a domain name to publishing the website. This solution is time efficient for the customer, as they do not allocate internal resources to such tasks and receive a ready-to-use product.
2. Use internal skills and set up the website on their own. This implies that members of the team have the skill and experience to create the website. The team has to allocate responsibilities, select and ensure a domain name, find and pay for a reliable host, use a template or create their own (based on their level of skills), insert and organize all content and then manage it.

Both solutions require a significant amount of money and, if the second option is selected, time and skills while at the same time, they completely depend upon the server and the data center that hosts them. In addition, there are fixed costs that need to be paid on an annual basis for maintaining it. Businesses may end up paying more than they actually use, limiting collaboration within the business teams due to teams operating in silos, easily maxing out their budget and facing scalability problems, which is one the biggest issues for companies when they plan for growth.

Cloud computing has been introduced as a new approach to satisfy customers' needs and is expected to continue its impressive growth. Cloud technologies facilitate data storage, data exchange and organization amongst businesses and individuals, provide flexibility, vastly reduce infrastructure costs and allow the workforce to better concentrate on their work, leading to increased productivity and efficiency.

However, the relationship between Cloud providers and companies/individuals is no different to any other model of service: supplier – buyer. This one-way model

allows customers to store, manage and share data privately or publicly, using the Cloud service that the supplier provides.

The business model Cloud computing is founded on is a very intriguing and value centered model as it enables users to pay per use instead of paying on a time-set basis, regardless of the use they make. Distributed Cloud technology introduces a new type of Cloud services based on peer-to-peer technology that aims to facilitate and enhance content creation and resource sharing. The idea of distributed Cloud computing is to combine the Grid and Cloud concepts. For a distributed Cloud, users' workstations provide their own computing recourses (storage and computing power) to Cloud participants. The network architecture is based on the principle of equal interaction between nodes. The number of users who may share the resource increases as a function of the resource's popularity. This means that the more popular a resource is, the more the users that can access it and share it. Apart from the above advantages, the unique point of differentiation lies within scalability. Peer-to-peer technology enables users to scale their content for free, regardless of the scalability level.

As a result, both, small and big companies can use the same technological basis to scale their content and grow to new levels.

4 Distributed P2P Based Cloud Service Model

The focus of the study is to improve the process of allocating resources between user nodes in a Cloud with a distributed architecture [2] and to reduce the response time for such a Cloud. Fig. 1 illustrates an implementation of the proposed approach.

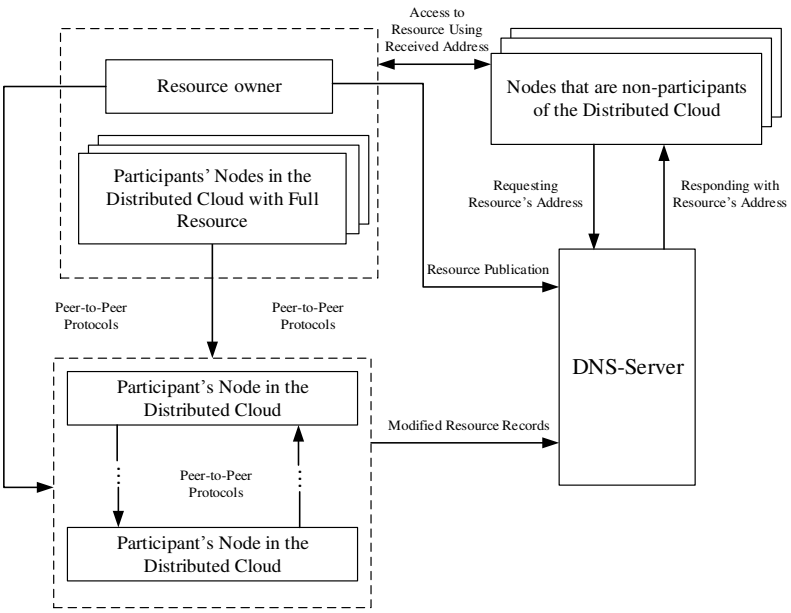


Fig. 1. Service model of the distributed Cloud

The solution can be found in organizing the decentralized resources between user nodes in a Cloud by using peer-to-peer protocols. Each node acts in 3 ways: as a client when making a request to the resource, as a server when responding to requests providing the resources and as a resource owner who has permission for full access. The initial stage entails the process of publishing the resource by its owner. It consists of creating an association between the domain name and the resource by adding the network address of the owner and the unique resource identifiers to DNS-record.

The domain name is used as a unique identifier while the resource allocation process is coordinated through the DNS-server by modifying the resource records. After the participant's node processes the request of a resource, it addresses it to the DNS-server and receives the address of the resource owner. The next stage is the resource replication process, which is implemented by a cache of request, and the response data on the storage of the user's workstation node makes it possible to share the result of the cache with other Cloud nodes. As a result, the node that responds to a request acts as a server and its address is added to one of the resource records of the DNS-server. It enables the node participant of the distributed Cloud to access the resource from several addresses contained in the DNS records, due to the fact that the requested resource is replicated, thus increasing the availability of the service. The more popular the resource, the more nodes can share it.

However, server functionality requires from the participant node to share the hardware resources of its workstation, such as storage space and processing power. The participants' nodes that interact with each other are equal and the implementation of the distributed Cloud on the users' side is achieved through the installation of a software on the workstations of each participant node. Furthermore, it is possible for participants' nodes that are non-members to access the resources through a regular request to DNS-server and to get the network addresses of the owner's station or the participants' nodes with the full resource. After receiving their addresses, they can interact with each other. However, it is important to note that for such users, the bandwidth is limited by the number of users that use the standard interaction mechanism.

5 Response time to a request

The response time to a remote Cloud server depends on several factors, such as customers' geographical location in relation to the server, the available bandwidth of communication channels and network interfaces, the number of concurrent user connections to the server, the rate of requests, the hardware configuration of the server etc. [11].

The average response time (receiving the service) for the end-user in a Cloud client-server architecture in general form can be expressed by the following formula:

$$t_{rrt\_c\text{-}s} = t_{base\_c\text{-}s} + t_{response\_serv} + t_{response\_BD}, \quad (1)$$

where $t_{rrt\_c\text{-}s}$ – the average time to access the resource,

$t_{base\_c\text{-}s}$ – the basic transmission delay on the communication channels,

$t_{response\_serv}$ – the server response time,

$t_{response\_BD}$ – the data base response time.

Consequently, the basic delay of the network transmission response is determined by the ratio of the data transmitted to the bandwidth:

$$t_{base\_c\text{-}s} = V/C, \tag{2}$$

where V – the data response size,

C – the available bandwidth.

The available user goodput is limited by the following: - the network adapter, - actual Internet access speed determined by the ISP, - the server bandwidth, - communication channels [12]. Thus, all users that are concurrently connected to the server evenly share the bandwidth. It means that the available goodput of the connection between the client and the server is determined by the goodput of the "bottle neck" of the route between the client and the server [13].

Therefore, to determine the value of the available goodput, the following expression can be used:

$$C = min\ (G, G_{serv}, G_{link}), \tag{3}$$

where G – the available user goodput,

G_{serv} – the available server goodput,

G_{link} – the available link goodput.

In cases where nodes of the distributed Cloud interact with each other, the average response time is defined by the sum of the basic time delay of the network component and the delay that is related to the process of finding and selecting the nodes that provide part of the resource, combining all the parts of the resource together and other time delays.

$$t_{rrt\_p2p} = t_{base\_p2p} + t_{interaction}, \tag{4}$$

where t_{rrt_p2p} – average time to access the resource,

t_{base_p2p} – basic transmission delay of the communication channels,

$t_{interaction}$ – interaction delay between nodes that provides resource or its part.

The basic transmission delay on communication channels, as previously mentioned, is determined by the bandwidth of the "bottle neck" of the network. The transfer of several parts of the resource may occur concurrently from multiple nodes with sufficient bandwidth. The overall delay network component will be determined by the slowest transmission time:

$$t_{base\_p2p} = max\ (t_{b\_p2p\_1}, t_{b\_p2p\_2}, \ldots, t_{b\_p2p\_N}), \tag{5}$$

where $t_{b\_p2p\_i}$ – the basic delay of the transmission communication channel from node i,

N – the number of nodes, from where receiving the resources occur concurrently.

Given the constraints of the available bandwidth, the user network adapter and the internet speed connection:

$$t_{b\_p2p\_i} = \begin{cases} \dfrac{V_i}{min\ \left(G,\ G_i,\ G_{link\_i}\right)}, \sum_{i=1}^{N} G_i \le G; \\ \dfrac{V}{G}, \sum_{i=1}^{N} G_i > G. \end{cases} \quad (6)$$

where G – the available user goodput,
G_i – the available goodput of i node,
G_{link_i} – the available link to the i node goodput,
V_i – the size of the resource V, provided by node i.

For a comparative evaluation of the average response time to a request for centralized client-server and distributed Cloud architectures, initial data (Table 2) is collected based on the analysis of the research [11-14].

Table 2. Initial data

Available user goodput, G	10 Mb/s
Available server goodput, G_{serv}	10 Gb/s
Available link goodput, G_{link}	8 Mb/s
Available goodput of i node, G_i	2 Mb/s
Available link to the i node goodput, G_{link_i}	5 Mb/s
The data response size, V	100 kB

Fig. 2 depicts the results of an estimation of the response time to a request for a centralized client-server *t_rrt_c-s* and a distributed *t_rrt_p2p* Cloud architecture for different numbers of concurrent users.

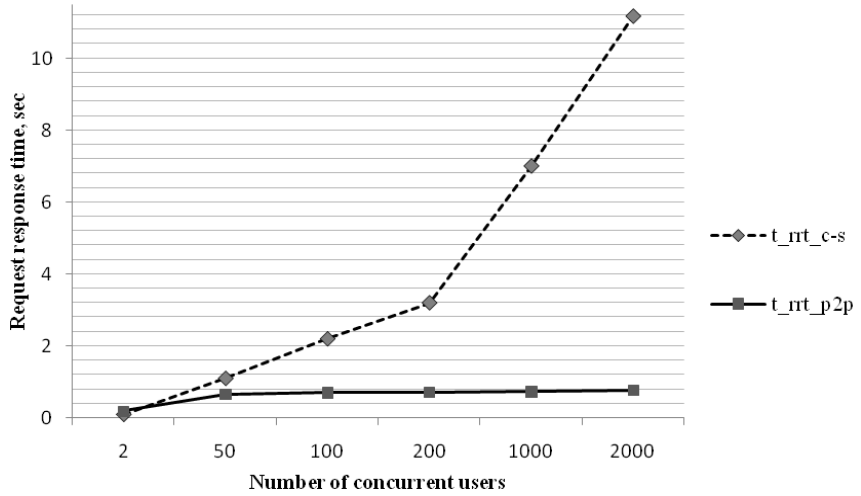


Fig. 2. Results of the response time estimation

Analyzing the graphical representation of the two types of architectures' behavior with the number of concurrent users makes it possible to conclude that the selected set

of input data can achieve a significant reduction of response time in the case of a distributed Cloud architecture with 15 or more concurrent users that all interact with each other.

6 Service Availability

As previously mentioned, the main advantage of a distributed cloud architecture is the high degree of resource replication. The copies of the resource are distributed among the nodes that had previously requested it. Thus, the number, states and hardware properties of the nodes will determine the Service Availability. There are different methods for the evaluation of Service Availability [15]. Within the scope of this study, the evaluation of Service Availability deployed on two types of cloud architectures is considered: centralized and distributed. Furthermore, Service Availability evaluation does not take into account the hardware, software and network failures. It is assumed that service is available when there are no performance-related failures that usually occur when incoming requests are not served due to limited capacity of the server. If the service is implemented based on the standard cloud server, then the probability that an arriving request is lost due to buffer overflow is described by the formula [15]:

$$P_b = \begin{cases} \rho^b \cdot \dfrac{1-\rho}{1-\rho^{b+1}}, \rho \neq 1; \\ \dfrac{1}{b+1}, \rho = 1. \end{cases}, \quad (7)$$

where ρ – the server load,

b – the server input buffer size.

The server's behavior can be modeled by a M/M/1/b queue. Then, when the steady state probability of the up state corresponds to the system's steady-state availability and when it is equal to 1, then the availability of the service is:

$$A_{(WS)} = (1 - P_b). \quad (8)$$

In the case of applying the distributed approach, the service is successfully provided as long as at least one of the replication nodes is available. The model of the system's behavior is described as M/M/c/b queue, where c - the number of available nodes that function as a replication, b – the node input buffer size. The probability of requests being lost due to buffer overflow is given by [15]:

$$L_{b(c)} = \begin{cases} \frac{\rho_n{}^{b_n}}{c^{b-c} \cdot c!} \cdot \left[\sum_{j=0}^{c-1} \frac{\rho_n{}^{j}}{j!} + \sum_{j=c}^{b} \frac{\rho_n{}^{b}}{c^{j-c} \cdot c!} \right]^{-1}, b_n \geq c; \\ \frac{\rho_n{}^{b_n}}{b_n!} \cdot \left[\sum_{j=0}^{b_n} \frac{\rho_n{}^{j}}{j!} \right]^{-1}, b_n < c. \end{cases} \quad (9)$$

where c – the number of replication nodes,
ρ_n – the node load,
b_n – the node input buffer size.
Similarly (2), the availability of the service is:

$$A_{(NS)} = (1 - L_{b(c)}). \quad (10)$$

The initial data that is used for the evaluation of Service Availability is taken from [14]. The input data and evaluation results are summarized in tabl. 3:

Table 3. Input data and estimation results

ρ	b	c	ρ_n	b_n	$A_{(WS)}$	$A_{(NS)}$
1	3000	100	1	7	0.99966	0.999927

As seen from the table, for the given set of input parameters, Service Availability is implemented through passive replication based on the above properties of the distributed cloud. Service Availability increases significantly with the number of replication nodes is increased and is, therefore, dependent on the popularity of the resource. However, in cases where the service is insufficiently popular and has a low degree of replication nodes, it would be more appropriate for the implementation to be based on a centralized cloud architecture, where the replication of existing nodes can reduce the load on the server. In order to implement an autonomous and stable operation of the service-based distributed cloud infrastructure without a server, it is important to determine the number of nodes' replication as sufficient enough. This is a crucial area for further research. Furthermore, it is necessary to consider a new business model for such an approach.

7 Moulding a new business model

Customers of distributed Cloud vary from individuals who plan to make a website for personal reasons, to freelancers or professionals who need a reliable service at a smart price, all the way to small and large corporations who seek to be innovative but without compromising their financial resources.

The use of websites is global but the needs are very different, depending on the quality, quantity, target group and nature of the information of it. Studying the market and creating a list of questions to guide the team throughout the process of identifying

each customer group, the following segments have been determined and are presented in table below.

Table 4. Customer segments and needs

Segment	Needs
Private individuals, blogs, small societies	- Interested in a small number of websites - Personal use mainly - Not significantly big amounts of data - Seek low prices and easy-to-use solutions
Professionals, freelancers, businesses	- Minimize IT costs - Flexibility & reliability - Scalability "Value for money" - Security
Startup companies and special organizations	- Low costs to create their web identity - Accessibility - Scalability - Use of innovative tools - Promotional tools

Emerging markets and technologies consist of a number of risks that should be taken into consideration before venturing. Distributed Cloud computing, slightly lagging Cloud computing, is at the beginning of its Life Cycle, where the early majority has already started adopting the technology for a number of daily applications. For a startup company, this point is a good one to enter the market, provided it can offer a unique differentiation and a well perceived value to its customers.

Various technologies have not only introduced new benefits and solutions to existing and new needs, but have also encouraged business models and strategies to change accordingly, in order to accommodate new trends and expectations. Cloud computing consists one of the most revolutionary technologies, mainly due to the fact that it shapes a different future. Through a shift in business conduct, it further empowers existing and new parties allowing more versatility, flexibility and innovation to grow.

However, technology does not create value on its own. It is the design and application of a sustainable and evolving business model that enables technology to create value for its users. A successful and well-developed business plan may result in cost reduction, strategic flexibility or even reduction in risk, amongst other benefits.

Existing business models have a distinct separation in roles. As depicted in Ostewalder's business model canvas, the company works with key partners and suppliers in order to create value for its customers and maintain a good and profitable relationship that will ensure a stable and increasing revenue stream. So far, usual business models may be characterized as "non-interactive" models, as the end-target (the customer) does not participate. Distributed Cloud enables business models to

change and include their customers in the value creation process. Fig. 3 depicts the distinct roles and interactions between participating sides within distributed Cloud, as structured in the business model canvas.

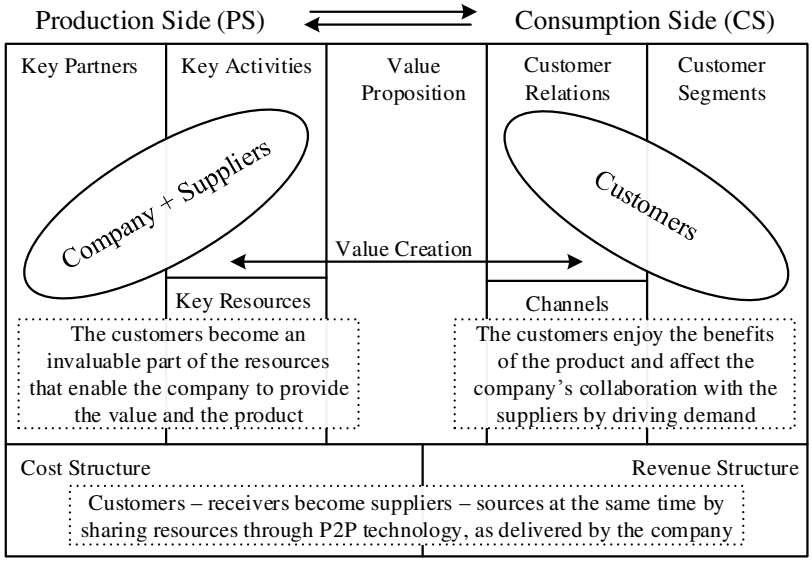


Fig. 3. Business model canvas

The Production Side (PS) refers to the cluster of parties that work together to create and deliver the value to the customers. Based on the customers' needs and demand, the PS organizes itself to ensure cost reduction, trusted relationships and quality partnerships that will lead to the product. The Consumption Side (CS) consists of the market cluster – the customers who are using the product and enjoy the benefits and the value it creates. Their needs and their demand is the major factor based on which the PS makes alterations and adjustments. For this reason, market and business intelligence is extremely critical, as the data provides companies with sufficient intel to allow trends predictions and needs insights. The capabilities of distributed Cloud technology set new foundations for business models to evolve and grow in order to better facilitate all interested parties. Distributed Cloud technology is a technology that interacts with users, depends on user popularity and constantly moves to adapt to its users. By natural consequence, business models that correspond to P2P technology need to interact with the user and adapt to changes in demand.

Distributed Cloud approach sets new roles and multiple sides to the business model canvas. Companies do not simply work with suppliers to create a product and customers and users of the technology do not simply purchase or enjoy the value of the product. Instead, the company becomes the technology facilitator and the customers become the suppliers as well. As a result, the business model that emerges is a model that is "alive" and "evolving" with change and a series of "participatory business models" is introduced, where customers obtain a double identity – that of the consumer and that of the source-supplier.

It is essential to notice that the Production Side of the business model canvas does not have any interaction with the Consumption Side, other than any alterations that result from business intelligence.

The double role of the customers is the key to this new business model ecosystem, significantly affecting the interaction between customer-user and customer-supplier, through the technology.

It is important to note that in such business models, monitoring intellectual property and rights is challenging and needs to be dealt with utmost care and responsibility in order to ensure protection of all participants.

In the field of distributed Cloud for website applications, the use of such a dynamic business model is essential, as it is the core of value creation and value delivery to customers. Users purchase or subscribe to the service to enjoy the benefits of the distributed Cloud solution as provided by the company, in the form of a software, thus becoming customers-consumers. In return, other customers-consumers that request access to the content of the website through the same software service, retrieve it from fellow customers-consumers who, having a virtual footprint of the information on their computers, they now become sources-suppliers for the new customers-consumers.

As a result, the group of people who share a common interest in the content they seek, participate in a "shared cluster" of information and act as both, sources and consumers of the content. Apart from the development of new business models, companies may find it beneficial to use distributed Cloud as it offers new options in terms of budget and scalability. Budget and scalability consist of the two main issues that companies face. Increased demand may indeed increase revenue stream. However, in order to meet this demand, companies need to invest in more resources. Even with current cloud services, suppliers' costs are high and not easy to handle.

Currently, companies increase their IT infrastructure and spend significant amounts on adding new serves in order to accommodate their customers' needs and the company's workload. Without this expansion, the company cannot achieve scalability.

Churn is also important, as users may activate and deactivate their nodes, altering the dynamic of the P2P system. However, the bigger the network of customers-sources, the more manageable is the churn, as the system will dynamically adapt to changes and maintain its efficiency.

Overall, distributed Cloud technology has a number of potential applications that can benefit different companies and users. However, the technology itself produces value when a major condition is met: the design of an appropriate, innovative and predictive business model that ensures value is captured and transferred to users, delivers results and revenue for the company and adapts to the dynamic nature of the system.

It is necessary to highlight that there is no correct business model. Different models may work as long as they incorporate the P2P values and focus on the double role of the customers. The term "participatory business models" is suggested to describe the new reality that companies and entrepreneurs are expected to face. Monetization through such a business model may be significantly more challenging compared to

more popular business models but this may set the foundations for a new way of conducting business and commercializing technologies, ideas and goods.

8 Case study

In order to implement distributed Cloud solution, a number of expenses categories has been created:

- Operational expenses refer to the expenses that the company needs to cover in order to operate and run smoothly. The expenses required for the running and operation are originally limited to the renting of a small server of 50-70 accounts and to the costs of premises and power (office rent, electricity etc.). The initial premises costs are small due to the fact that a big part of the work completed will be through computers and virtual desktops.
- Labour expenses refer to the amount of money the company needs to pay for the services of its human resources. In this category, all expenses that refer to salaries of regular employees or outsourced partners are included.
- Variable expenses refer to the expenses of other parameters such as marketing campaigns, royalties to the university, depreciation of the investment within maximum 2 years etc. Variable expenses include marketing budget, university royalties of 3% for the first year and 5% onwards and depreciation of the investment cost within 2 years.

The proposed solution has a number of implementation stages in order to be fully developed and be ready to use. Strategic planning before starting the development will ensure time efficiency, productive allocation of tasks and smart use of resources. More specifically, the development stages of the project are summarized in the table below:

Table 5. Suggested implementation stages

Stage	*Description*
Stage 1. Implementation of the PaaS (platform as a service) service model, based on the distribution of tasks, services, websites and storage between the users.	Within the first stage, a number of steps are included: implementation of the static website functionality, distribution of the website tasks and storage between the users' hardware, and finally, implementation of the general service.
Stage 2. Implementation of the IaaS (infrastructure as a service) service model.	Implementation of the special software layer, which is responsible for distributing the system requests of the guest operating system between the participants' devices, allowing the running of a virtual machine.
Stage 3. Implementation of the SaaS (software as a service) service model.	The last stage assumes improvement of the software based on the specifically configured virtual machine that was mentioned above.

The distributed Cloud solution can begin its commercialization from within the academic society. The educational sector is a potential big customer that could highly benefit from the proposed approach, and the use of academic and EU connections can greatly help spread the technology. The aim is to begin locally by promoting the proposed solution through academic events and contacts with the respective IT administration departments in order to expand to more universities locally and internationally.

9 Conclusion

Cloud Computing has arrived at a very good timing to introduce a new reality based on current needs and future aspirations. Granted, as a new technology, it requires a number of years to set and to become popular, constantly raising awareness amongst people and introducing them to its benefits. The above trends demonstrate the high prospects of different types of Cloud Computing that is expected to grow in the following years in many markets and to extend to a variety of applications. Competition is an expected to be high and technology progress will demand constant update of versions and improvement of products and solutions, in order to successfully supply an ever growing market and a rapidly changing business and social environment.

The Distributed P2P based Cloud is a new solution that aspires to change the way companies and organizations work with regards to creating, publishing and sharing content. The purpose of the Distributed Cloud approach is to provide P2P Cloud services to individuals, companies and organizations with the view to facilitating cost-effective scalability, flexibility and efficiency and enhancing the experience of creating, organizing, publishing and sharing content. Its competitive advantage lies within the concept of reliability and scalability at significantly lower costs, allowing customers to differently allocate their financial resources or to grow even on a budget.

In addition, well-structured, targeted and smart marketing strategies are necessary to be developed in order to ensure strategic growth of the business sector and brand awareness amongst existing and prospective customers. The proposed solution combines Cloud computing and Grid technology with peer-to-peer networks through a software that allows users to participate in a single, decentralized Cloud system and use their workstations to allocate network resources. As a result, it partially or completely eliminates the need to use powerful, high-performance servers in virtual data centers and, ultimately, reduces energy consumption and negative impacts on the environment.

By applying Distributed Cloud technology, the dynamics of the system changes so that users become sources and consumers at the same time. This is the very core value of this technology as it enhances cost effective scalability and changes the way business is conducted, through a dynamic, alive and self-adjusting business model.

The interchanging roles of customers and suppliers within such a participatory business model encourages companies and entrepreneurs to focus more strongly on the value that can be obtained through the P2P technology within the distributed

cloud, rather than the marketing of the product itself. Consumers are becoming more and more informed about how technology works and what benefits each provider gives them. This means, that shifting towards value creation and focus through customer participation may actually help companies differentiate themselves from the mass, attract more customers and, eventually, contribute to a new corporative culture.

References

1. GetSync, https://www.getsync.com/
2. Yanovskaya, O., Yanovsky, M., Kharchenko, V.: The Concept of Green Cloud Infrastructure Based on Distributed Computing and Hardware Accelerator within FPGA as a Service. In: Design & Test Symposium (EWDTS), pp. 45–48. IEEE Press, Kyiv(2014)
3. Cisco Global Cloud Index: Forecast and Methodology, 2013–2018, http://www.cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index-gci/Cloud_Index_White_Paper.pdf
4. Xu, H., Li, B.: A Study of Pricing for Cloud Resources. SIGMETRICS Perform. Eval. Rev. 40, 3–12, New York, NY (2013)
5. Spacemonkey, https://www.spacemonkey.com/
6. Project Maelstrom, http://project-maelstrom.bittorrent.com/
7. Wuala, http://wuala.com/
8. Sherly, https://sher.ly/
9. Symform, http://www.symform.com/
10. Berkeley Open Infrastructure for Network Computing (BOINC), https://boinc.berkeley.edu/
11. Martinello, M.: Availability Modeling and Evaluation of Web-based Services-A pragmatic approach. (2005)
12. Gorbenko, A., Kharchenko, V., Mamutov, S., Tarasyuk, O., Romanovsky, A.: Exploring Uncertainty of Delays as a Factor in End-to-End Cloud Response Time. In: Ninth European Dependable Computing Conference (EDCC), pp. 185–190. IEEE Press, Sibiu (2012)
13. Gorbenko, A., Romanovsky, A.: Time-Outing Internet Services. Security & Privacy, IEEE, 11(2), 68–71. doi: 10.1109/MSP.2013.43 (2013)
14. Elyasi-Komari, I., Gorbenko, A., Kharchenko, V.S., & Mamalis, A.: Analysis of Computer Network Reliability and Criticality: Technique and Features. IJCNS, 4(11), 720–726. doi: 10.4236/ijcns.2011.411088 (2011)
15. Benchmark Results, http://docs.oracle.com/cd/E13218_01/wlp/docs81/capacityplanning/capacityplanning.html
16. Dabek, F., Li, J., Sit, E., Robertson, J., Kaashoek, M.F., Morris, R.: Designing a DHT for Low Latency and High Throughput. In: Conference on Symposium on Networked Systems Design and Implementation (NSDI), vol. 1, pp. 85–98. USENIX Association Berkeley, San Francisco (2004)
17. Sacramento, V., Endler, M., Souza, C.D.: A Privacy Service for Location-Based Collaboration among Mobile Users. Journal of the Brazilian Computer Society 14, 41–57 (2008)
18. Olshefski, D.P., Nieh, J., Nahum, E.: Ksniffer: Determining the Remote Client Perceived Response Time from Live Packet Streams. In: 6th conference on Symposium on Operating Systems Design & Implementation, vol. 6, pp. 333–346. USENIX Association Berkeley, San Francisco, (2004)

Automated Development of Markovian Chains for Fault-Tolerant Computer-Based Systems with Version-Structure Redundancy

BogdanVolochiy[1], Oleksandr Mulyak[2], Vyacheslav Kharchenko[3]

[1]National University Lviv Polytechnic, Lviv, Ukraine
bvolochiy@ukr.net
[2]RPC "PromTechnoServis Ukraine", Kyiv, Ukraine
mulyak@prom-technoservice.com
[3] National Aerospace University "KhAI", Kharkiv, Ukraine
v.kharchenko@khai.edu

Abstract. Reliability design of fault-tolerant computer-based systems with version-structural redundancy and multiply software updates involves solving number of issues. This paper outlines an availability model of the computer-based systems which shows the algorithm for reliability behavior. For various configurations of the computer-based systems, the use of the proposed model and problem-oriented software, ASNA represents the ability to automate constructed the Markovian chains. This model includes a number of settings: failure rate of the software; numbers of software updates; duration of software updates; the structure of the system's hardware and reliability indicators. The proposed model for the automated development of Markovian chains is subject to the adaptation of the structure of the hardware of computer-based systems and/or the algorithms of reliability behavior. This allows us to obtain a new model and the feasibility to automate development of the Markovian chains.

Keywords. Markovian Chains, Automated Reliability Design, Fault-Tolerance System, Version-Structural Redundancy, Common Sliding Standby, Hot Standby, Cold Standby

Key Terms. Mathematical Modeling, Method, Software Systems.

1 Introduction

1.1 Motivation

Nowadays the developments of fault-tolerant computer-based systems (FTCSs) are a part of weaponry components, space, aviation, energy and other critical systems. One

of the main tasks is to provide requirements of reliability, availability and functional safety. Thus the two types of possible risks relate to the assessment of risk, and to ensuring their safety and security.

Reliability (dependability) related design (RRD) [1-6] is a main part of development of complex fault-tolerant systems based on computers, software (SW) and hardware (HW) components. The goal of RRD is to develop the structure of FTCS tolerating HW physical failure and SW designs faults and assure required values of reliability, availability and other dependability attributes. To ensure fault-tolerance software, two or more versions of software (developed by different developers, using other languages and technologies, etc) are used [7]. Therefore use of structural redundancy for FTCS with multiple versions of software is mandatory. When commissioning software some bugs (design faults) remain in its code [8], this leads to the shut-down of the FTCS. After detection the bugs, a software update is carried out. These factors have influence on the availability of the FTCS and should be taken into account in the availability indexes. During the operation of FTCS it is also possible that the HW will fail leading to failure of the software. To recover the software operability, an automatic restart procedure, which is time consuming, is performed. The efficiency of fault-tolerant hardware of FTCS is provided by maintenance and repair.

Insufficient level of adequacy of the availability models of FTCS leads either to additional costs (while underestimating of the indexes), or to the risk of total failure (when inflating their values), namely accidents, material damage and even loss of life. Reliability and safety are assured by using (selection and development) fault-tolerant structures at RRD of the FTCS, and identifying and implementing strategies for maintenance. Adoption of wrong decisions at this stage leads to similar risks.

1.2 Related Works Analysis

Research papers, which focus on RRD, consider models of the FTCS. Most models are primarily developed to identify the impact of one the above-listed factors on reliability indexes. The rest of the factors are overlooked. Papers [4, 5] describe the reliability model of FTCS which illustrates separate HW and SW failures. Paper [6] offer reliability model of a fault-tolerant system, in which HW and SW failures are differentiated and after corrections in the program code the software failure rate is accounted for. Paper [8] describes the reliability model of the FTCS, which accounts for the software updates. In paper [10] the author outlines the relevance of the estimation of the reliability indexes of FTCS considering the failure of SW and recommends a method for their determination. Such reliability models of the FTCS produce analysis of its conditions under the failure of SW. This research suggests that $MTTF_{system}=MTTF_{software}$. Thus, it is possible to conclude that the author considers the HW of the FTCS as absolutely reliable. Such condition reduces the credibility of the result, especially when the reliability of the HW is commensurable to the reliability of the SW. Paper [11] presents the assessment of reliability parameters of FTCS through modelling behavior using Markovian chains, which account for multiple software updates. Nevertheless there was no evidence of the quantitative assessments of the reliability measures of presented FTCS.

In paper [12], the authors propose a model of FTCS using Macro-Markovian chains, where the software failure rate, duration of software verification, failure rate and repair rate of HW are accounted for. The presented method of Macro-Markovian chains modelling [12, 13] is based on logical analysis and cannot be used for profound configurations of FTCS due to their complexity and high probability of the occurrence of mistakes. Also there is a discussion around the definition of requirements for operational verification of software of the space system, together with the research model of the object for availability evaluation and scenarios preference. It is noted that over the last ten years out of 27% of space devices failures, which were fatal or such that restricted their use, 6% were associated with HW failure and 21% with SW failure.

Software updates are necessary due to the fact that at the point of SW commissioning they may contain a number of undetected faults, which can lead to critical failures of the FTCS. Presence of HW faults relates to the complexity of the system, and failure to conduct overall testing, as such testing is time consuming and needs substation financial support. To predict the number of SW faults at the time of its commissioning various models can be used, one for example is Jelinski-Moranda [14].

A goal of the paper is to suggest a technique to develop a Markovian chain for complex FTCS with different redundancy types (first of all, structure and version) using the proposed formal procedure and tool. The main idea is to decrease risks of errors during development of MC for systems with very large (tens and hundreds) number of states. We propose a special notation which allows supporting development chain step by step and designing final MC using software tools. The paper is structured in the following way. The aim of this research is calculating the availability function of FTCS with version-structural redundancy and double software updates.

To achieve this goal we propose a newly designed reliability model of FTCS. As an example a special computer-based system of space radio-technical complex is researched (Fig.1). The following factors are accounted for in this model: overall reserve of FTCS and joint sliding reserve of modules of main and diverse FTCS; the existence of two software versions; SW double update; and automatic software reboot, if its failure was caused by the HW physicals fault.

Structure of the paper is the following. Researched FTCS is described in the second section. An approach to developing mathematical model based on Markovian chain and detailed procedure for the FTCS are suggested in the third and fourth sections correspondingly. Simulation results for researched Markov's model are analyzed in the section 5. Last section concludes the paper and presents some directions of future researches and developments.

2 Researched fault-tolerant computer based system with structure-version redundancy

The researched FTCS with structure-version redundancy is shown on figure 1. To ensure the minimal FTCS downtime, overall hot standby with other version of software is used.

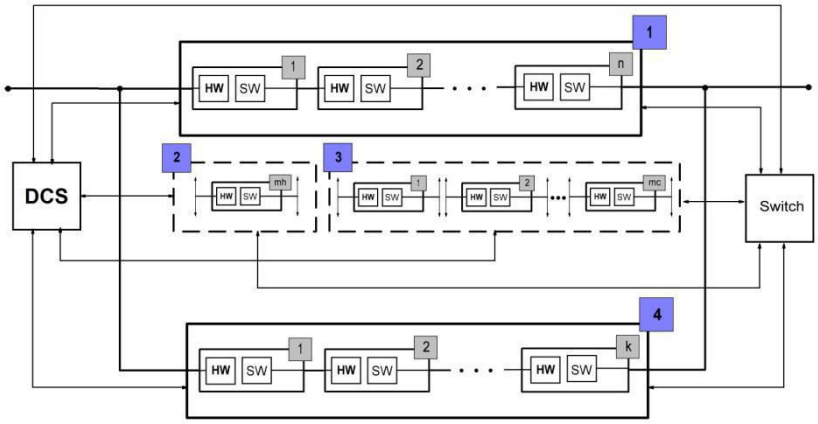


Fig. 1. Fault-tolerant Computer Based System (1 – main system, 2 – hot standby, 3 – cold standby, 4 – diverse system, DCS – Diagnostics Control System).

The FTCS consists of: a main system comprising modules; diverse system consist of k - modules; for two systems, the common sliding standby of modules is envisaged, the first module in hot standby and other in cold standby; a diagnostics control system determines the state of HW and SW, and manages the redundancy; and a switch is connected the modules to the main and diverse systems.

3 An approach to developing an availability model for FTCS with software update and restarting

An approach to the development of availability model for FTCS with double software updates and automatically software restart in the form of Markovian chains is presented in figure 2. During the operation of computer based system there are the following states: S1, S4 and S7 – system operable states; S2, S5, S9, S10 –inoperable states, in which SW updates, are conducted; S3, S6, S8 – inoperable states in which software restart after physical failure is automatically conducted; S11, S12, S13 – inoperable states in which HW is repaired after physical failure.

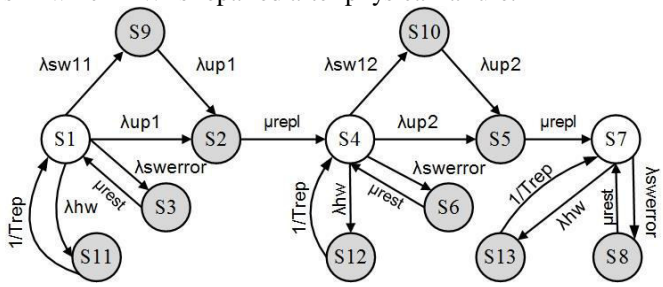


Fig. 2. Markovian chain which show the reliability behavior of computer-based system with double software updates and automatic restart

The system functioning after state S1 can unfold in four possible ways: the system moves to state S9 with rate λ_{sw11} after failure of first software version; the system move to state S3 with rate $\lambda_{swerror}$, after temporary failure of SW; the system move to state S2 with rate $\lambda_{up1}=1/T_{up1}$ (T_{up1} – duration of bugs correction in software) after finished the first version software operation (ready to use second version of software); the system move to state S11 with rate λ_{hw}, after physical failure.

The system functioning after state S4 can unfold in four possible ways: the system move to state S10 with rate λ_{sw12} after failure of second SW version; the system move to state S6 with rate $\lambda_{swerror}$, after temporary failure of SW; the system move to state S5 with rate $\lambda_{up2}=1/T_{up2}$ (T_{up2} – duration of bugs correction in software) after finished the second version SW operation (ready to use the third version of SW); the system move to state S12 with rate λ_{hw}, after physical failure. The system functioning after state S7 can unfold in two possible ways: the system move to state S6 with rate $\lambda_{swerror}$, after temporary failure of SW; the system move to state S13 with rate λ_{hw}, after physical failure. System moves from state S3and S6 to states S1 and S4 with a rate of $\mu_{rest}=1/T_{rest}$ (T_{rest} – duration of SW restart).

When the system is in state S2 and S5, it replaced the version of SW with rate $\mu_{repl}=1/T_{repl}$ (T_{repl} – duration of software replacement). The system moves from states S9 and S10 to states S2 and S5 with rate $\lambda_{up1}=1/T_{up1}$ (T_{up1} – duration of bugs correction in software) and $\lambda_{up2}=1/T_{up2}$ (T_{up2} – duration of bugs correction in software).

After commissioning of the computer based system starts debugging the software and develops first update that takes time T_{up1}. Second update takes timeT_{up2} and involves finding all bugs in SW. Therefore after first SW update, the numbers of bugs decreases, and debugging software is more complex, and the development of second update takes more time $T_{up2}>T_{up1}$.

The described approach to the development of availability model for FTCS with software update and restart is used to build availability model of FTCS showed in figure 1.

4 Markov's model for FTCS with software update and restarting

The method of development Markovian chain of the FTCS is described in the monograph [9]. It involves a formalized representation of the object of study as a "structural-automated model". To develop this availability model of the FTCS one needs to perform the following tasks: develop a verbal description of the research object (fig. 1); define the basic events; define the component vector of states, which can be described as a state of random time; define the parameters for the object of research, which should be in the model; and shape the tree of the modification of the rules and component of the vector of states.

4.1 The procedures to describe behavior of the FTCS

The FTTS behavior is described by the following procedures.

Procedure 1. Detection of failure of the FTCS (hardware failure, software failure, temporary failure). Failure can occur in the main and diverse system.

Procedure 2. Detection of failure in the main or diverse subsystems of the FTCS.

Procedure 3. Connection of the module from hot standby to faulty subsystem.

Procedure 4. Connection of the module from hot standby to cold standby.

Procedure 5. Loading the software on the module with connections from cold to hot standby.

Procedure 6. Software restart.

Procedure 7. Development the software updates.

Procedure 8. Repair (replacement) of the HW of the FTCS.

4.2 A set of the events for the FTCS

According to described procedures which determine the behavior of FTCS, a list of events is composed. Events are presented in pairs corresponding to the start and the end of time intervals to perform each procedure. From this list of events for "structural-automated model" basic events are selected [9].

As a result of analysis, twelve basic events in particular were determined: **Event 1** – "Hardware failure of main system module"; **Event 2** – "Software failure of the main system module"; **Event 3 –** "Software fault of the main system module"; **Event 4** – "Hardware failure of the diverse system module"; **Event 5 –** "Software failure of the diverse system module"**; Event 6 –** "Software fault of the diverse system module"; **Event 7 –** "Module failure in hot standby"; **Event 8 –** "Termination of the procedure of the hot standby module connection to non-operational system"; **Event 9 –** "Termination of the procedure of the cold standby module transfer to non-operational system"; **Event 10 –** "Termination of the procedure of software reloading on the module with failure feature in its software work"; **Event 11 –** "Termination of the procedure of SW version renovation"; **Event 12 –** "Termination of the procedure of the HW repair".

4.3 Components of vector states for the FTCS

Components of the vector state that can also be described as a state of random time. To describe the state of the system, eleven components are used: V1 – displays the current number of modules in the main system (the initial value of components V1 equal to n); V2 – displays the current number of modules in the diverse system (the initial value of components V2 equal to k); V3 – displays the current number of modules in hot standby (the initial value of components V3 equal to m_h); V4 – displays the current number of modules in cold standby (the initial value of components V4 equal to m_c); V5 – displays which software version is operated by the main system (V5=0 – first version, V5=1 – second version, V5=2 – third version); V6 – displays which software version operated by diverse system (V6=0 – first version, V6=1 – second version, V6=2 – third version); V7 – displays the temporary SW failure in the main system; V8 – displays the temporary SW failure in the diverse system; V9 – displays the SW fault in the main system; V10 – displays the SW fault in the diverse system; V11 – displays the number of non-operational units, due to HW fault.

4.4 The parameters of the FTCS Markov's model

Developing Markov's model of the FTCS, its composition and separate components should be set to relevant parameters in particular: n – number of modules that are the part of the main system; k – number of modules that are the part of the diverse system; m_h –number of the modules in the hot standby; m_c – number of the modules in the cold standby; λ_{hw} – the failure rate that is in main (diverse) system and in the hot standby; λ_{sw11}, λ_{sw12} – the failure rate of first and second software versions; $\lambda_{swerror}$ – the temporary failure rate of software; T_{up1}, T_{up2} – duration of the first and second software updates; T_{rest} – duration of software restart on the module; T_{switch} – duration of the module connections of the slight standby; T_{rep}– hardware repair duration.

4.5 Model of the FTCS for the automated development of the Markovian chain with software update and restart

According to the technology of analytical modeling, the discrete-continuous stochastic systems [9] based on certain events using the component vector state and the parameters that describe FTCS, and model of the FTCS for automated development of the Markovian chains are presented on the table 1.

Table 1. Model "Structural-Automated Model" of the FTCS for the automated development of the Markovian chains

Terms and conditions	Formula used for the intensity of the events	Rule of modification component for the state vector
Event 1. Hardware failure of main system module		
(V1=n) AND (V3>0) AND (V9=0)	$V1\cdot\lambda_{hw}$	V1:=n; V3:=V3-1; V11:=V11+1
(V1=n) AND (V3=0) AND (V9=0)	$V1\cdot\lambda_{hw}$	V1:=V1-1; V11:=V11+1|
Event 2. Software failure of the main system module		
(V1=n) AND (V3=0) AND (V9=0)	$V1\cdot\lambda_{swerror}$	V1:=V1-1; V7:=V7+1
(V1=n) AND (V3>0) AND (V9=0)	$V1\cdot\lambda_{swerror}$	V1:=n; V3:=V3-1; V7:=V7+1
Event 3. Software fault of the main system module		
(V1=n) AND (V5=0) AND (V9=0)	$V1\cdot\lambda_{sw11}$	V1:=V1-1; V5:=0; V9:=1
(V1=n) AND (V5=1) AND (V9=0)	$V1\cdot\lambda_{sw12}$	V1:=V1-1; V5:=1; V9:=1
Event 4. Hardware failure of the diverse system module		
(V2=k) AND (V3>0) AND (V10=0)	$V2\cdot\lambda_{hw}$	V2:=k; V3:=V3-1; V11:=V11+1
(V2=k) AND (V3=0) AND (V10=0)	$V2\cdot\lambda_{hw}$	V2:=V2-1; V11:=V11+1
Event 5. Software failure of the diverse system module		
(V2=k) AND (V3=0) AND (V10=0)	$V2\cdot\lambda_{swerror}$	V2:=V2-1; V8:=V8+1
(V2=k) AND (V3>0)	$V2\cdot\lambda_{swerror}$	V2:=k; V3:=V3-1; V8:=V8+1
Event 6. Software fault of the diverse system module		

(V2=k) AND (V6=0) AND (V10=0)	$V2 \cdot \lambda_{sw11}$	V2:=V2-1; V6:=0; V10:=1
(V2=k) AND (V6=1) AND (V10=0)	$V2 \cdot \lambda_{sw12}$	V2:=V2-1; V6:=1; V10:=1
Event 7. Module failure that is in the hot standby		
(V3>0) AND ((V9=0) OR (V10=0))	$V3 \cdot \lambda_{hw}$	V3:=V3-1; V11:=V11+1
Event 8. Termination of the procedure of the hot standby module connection to non-operational system		
(V1<n) AND (V3>0) AND (V11>0)	$1/T_{switch}$	V1:=V1+1; V3:=V3-1
(V2<k) AND (V3>0) AND (V11>0)	$1/T_{switch}$	V2:=V2+1; V3:=V3-1
Event 9. Termination of the procedure of the cold standby module transfer to non-operational CS		
(V3<mh) AND (V4>0)	$1/T_{switch}$	V3:=V3+1; V4:=V4-1
Event 10. Termination of the procedure of software reloading on the module with failure feature in its software work		
(V1<n) AND (V7>0)	$1/T_{rest}$	V1:=V1+1; V7:=V7-1
(V2<k) AND (V8>0)	$1/T_{rest}$	V2:=V2+1; V8:=V8-1
Event 11. Termination of the procedure of software version renovation		
(V1<n) AND (V5=0) AND (V9=1)	$1/T_{up1}$	V1:=n; V5:=1; V9:=0
(V1<n) AND (V5=1) AND (V9=1)	$1/T_{up2}$	V1:=n; V5:=2; V9:=0
(V2<k) AND (V6=0) AND (V10=1)	$1/T_{up1}$	V2:=k; V6:=1; V10:=0
(V2<k) AND (V6=1) AND (V10=1)	$1/T_{up2}$	V2:=k; V6:=2; V10:=0
Event 12. Termination of the procedure of the hardware repair		
(V1<n) AND (V2<k) AND (V11=2)	$1/T_{rep}$	V1:=n; V2:=k; V11:=0

As shown in Table 1, the model can be easily adapted for other fault-tolerant hardware configurations, or any number of software updates can be provided. For example, if the main and diverse system built with rule of voting 2-out-of-3, the description of the ***terms and conditions*** for the ***events of 1-6*** should be replaced with (V1=n) → (V1<=2) i (V2=k) → (V2<=2). In this way, we obtain the FTCSC for the automated development of the Markovian chains in which permanent and diverse fault-tolerant systems are built with rule of voting 2-of-3.

The number of software updates can be also changed. It is necessary to change vectors V5 and V6 the ***event 11***, that are responsible for the number of updates. For example, if there are three software updates, the entry component of the event will be as follows:

(V1<n) AND **(V5=2)** AND (V9=1)	$\mathbf{1/T_{up3}}$	V1:=n; **V5:=3;** V9:=0
(V2<k) AND **(V6=2)** AND (V10=1)	$\mathbf{1/T_{up3}}$	V2:=k; **V6:=3;** V10:=0

In order to present the necessary strategy of hardware repair can be transformed ***the event 12***.

4.6 Automated development of the Markovian chain and determining of availability function

The developed availability model of the FTCS gives the possibilities according to technology [9] for automated construct of the Markovian chains. This construction provides a software module ASNA [15]. The Markovian chains which take into account the following settings FTCS: n=2; k=2;m_h=0; m_c=0; λ_{hw}; λ_{sw11}, λ_{sw12}; $\lambda_{swerror}$; T_{up1}, T_{up2}; T_{rest}; T_{switch}; T_{rep}, are presented in figure 3. Information is available on the status of each software module ASNA we have on file "vector.vs", which is written in the form:

State 1: V1=2; V2=2; V3=0; V4=0; V5=0; V6=0; V7=0; V8=0; V9=0; V10=0; V11=0
State 2: V1=1; V2=2; V3=0; V4=0; V5=0; V6=0; V7=0; V8=0; V9=0; V10=0; V11=1
…….
State 121: V1=1; V2=1; V3=0; V4=0; V5=2; V6=2; V7=1; V8=1; V9=0; V10=0; V11=0

As the configurations of researched FTCS changes, the dimension of graphs increases. Therefore for the configuration of FTCS (Fig. 1) with one module in a hot reserve graph has 395 states and 976 transitions.

The proposed availability model of FTCS can be easily transformed for other features of the object of study. It is enough to: add / remove basic event (4.2); attach / remove components of the state vector (4.3); and include / exclude parameters that describe the studied system (4.4). Based on information about the work of FTCS an appropriate change in the model could be made (Fig 1).

Basing on the Markovian chains (Fig 3) formulas for designing of availability FTCS can be assembled. One measure of the availability of recovered FTCS reveals it is an availability function. Availability function of FTCS is calculated as the sum of the probability functions staying in operable states of chains. Basing on these states the FTCS availability function with parameters of FTCS n=2; k=2; m_h=0; m_c=0 is determined by the formula (1):

10

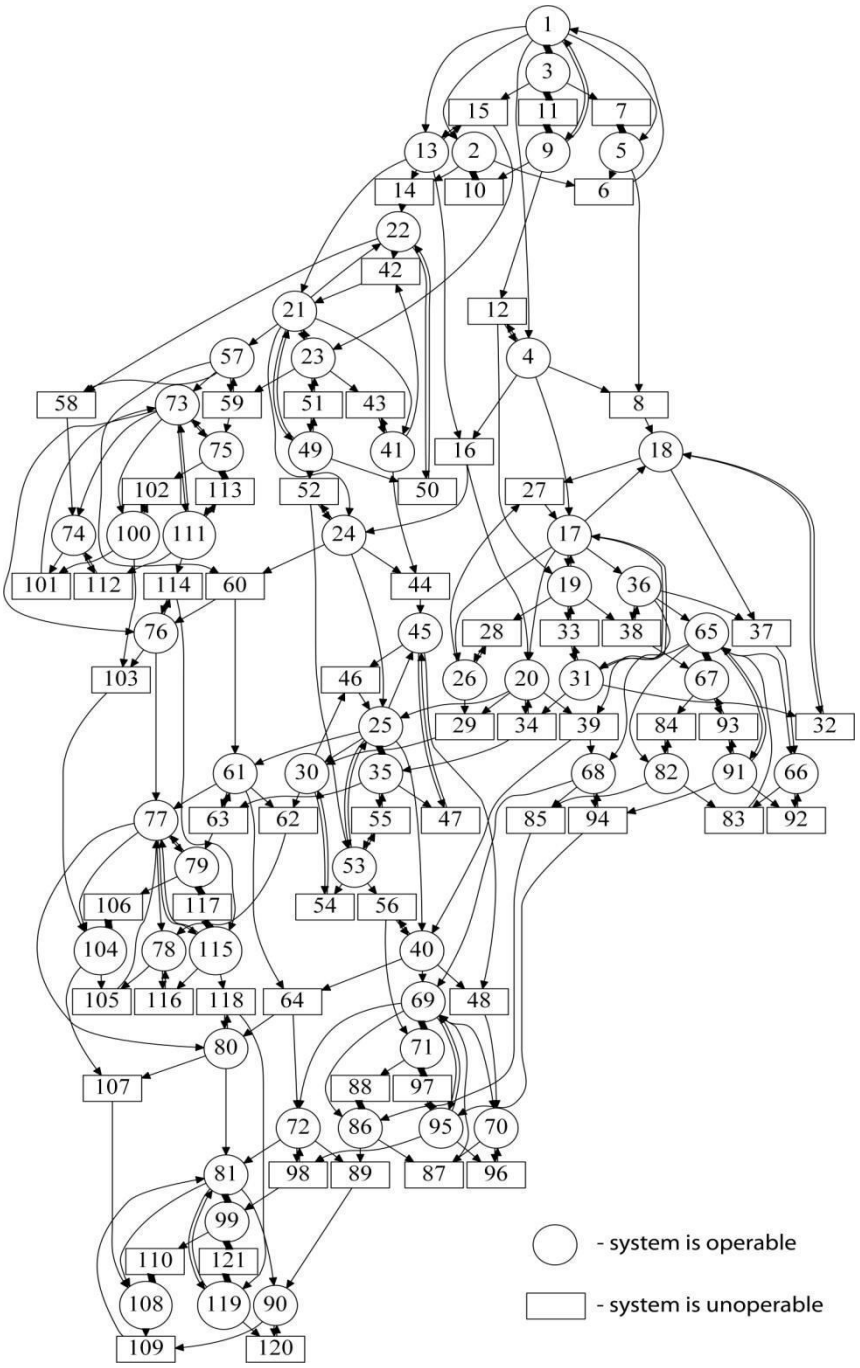


Fig. 3. The Markovian chains of the researched FTCS

$$K_{\Gamma}(t)=\sum_{i=1}^{5}P_i(t)+P_9(t)+P_{13}(t)+\sum_{i=17}^{26}P_i(t)+\sum_{i=30}^{31}P_i(t)+\sum_{i=33}^{36}P_i(t)+\sum_{i=40}^{41}P_i(t)+P_{45}(t)+P_{49}+$$
$$+(t)+P_{53}(t)+P_{57}(t)+P_{61}(t)+\sum_{i=65}^{82}P_i(t)+P_{86}(t)+\sum_{i=90}^{91}P_i(t)+P_{95}(t)+P_{95}(t)+\sum_{i=99}^{100}P_i(t)+ \quad (1)$$
$$+P_{104}(t)+P_{108}(t)+P_{111}(t)+P_{115}(t)+P_{119}(t)$$

Based on the Markovian chains (Fig.3) a system of differential equations (2) was formed. Its solution allows us to estimate the function availability value of researched FTCS.

$$\left.\begin{aligned}
&\frac{dP_1(t)}{dt}=-2\cdot\lambda_{hw}\left(P_2(t)+P_5(t)\right)-2\cdot\lambda_{sw11}\left(P_4(t)+P_{13}(t)\right)-2\cdot\lambda_{swerror}\cdot P_3(t)-\\
&\qquad-2\cdot\lambda_{swerror}\cdot P_9(t)+\frac{1}{T_{rep}}\cdot P_6(t)+\frac{1}{T_{rest}}\cdot\left(P_3(t)+P_3(t)\right)\\
&\frac{dP_2(t)}{dt}=2\cdot\lambda_{hw}P_1(t)-\frac{1}{T_{rest}}\cdot P_{10}(t)-2\cdot\lambda_{hw}P_6(t)-2\cdot\lambda_{swerror}\cdot P_{10}(t)-\\
&\qquad-2\cdot\lambda_{sw11}P_{14}(t)\\
&\frac{dP_3(t)}{dt}=2\cdot\lambda_{hw}P_1(t)+\frac{1}{T_{rest}}\cdot P_3(t)-2\cdot\lambda_{hw}P_7(t)-2\cdot\lambda_{swerror}\cdot P_{11}(t)-\\
&\qquad-\frac{1}{T_{rest}}\cdot P_{11}(t)-2\cdot\lambda_{sw11}P_{15}(t)\\
&\vdots\\
&\frac{dP_{121}(t)}{dt}=-\frac{1}{T_{rest}}\cdot P_{90}(t)+2\cdot\lambda_{swerror}\cdot P_{90}(t)+2\cdot\lambda_{hw}P_{119}(t)
\end{aligned}\right\}(2)$$

Initial conditions for the system (2) is $P_1(t)=1;\ \ P_2(t)...P_{121}(t)=0.$

5 Simulation results

5.1 Research of influence of software updates duration on the availability function

With the assistance of the proposed model, the following questions can be answered: What are the duration values of the first and the second software update (ensuring the values of the availability function of FTCS of the initial phase of its operation do not reach below the specified level)? What are the allowed duration values of the first and the second SW updates? How does the correlation between the first and the second SW updates influence on the availability function?

The first experiment is conducted for the condition where the duration of the first software update is significantly shorter than the duration of the second update. The duration of the first update is given within 10 - 50 hours, and the duration of the second update - 200 hours. The experiment is conducted with the following parameters FTCS: $\lambda_{hw}=1\cdot10^{-5}$ hour^{-1}; $\lambda_{sw11}=2\cdot10^{-3}$ hour^{-1}, $\lambda_{sw12}=1\cdot10^{-3}$ hour^{-1}; $\lambda_{swerror}=1\cdot10^{-2}$ hour^{-1}; T_{rest}=6 min; T_{switch}; T_{rep}=200 hour; T_{up2}=200 hour; (*line 1* -T_{up1}=10 hour; *line2* - T_{up1}=20 hour; *line3* - T_{up1}=30 hour; *line4* - T_{up1}=40 hour; *line5* -T_{up1}=50 hour).

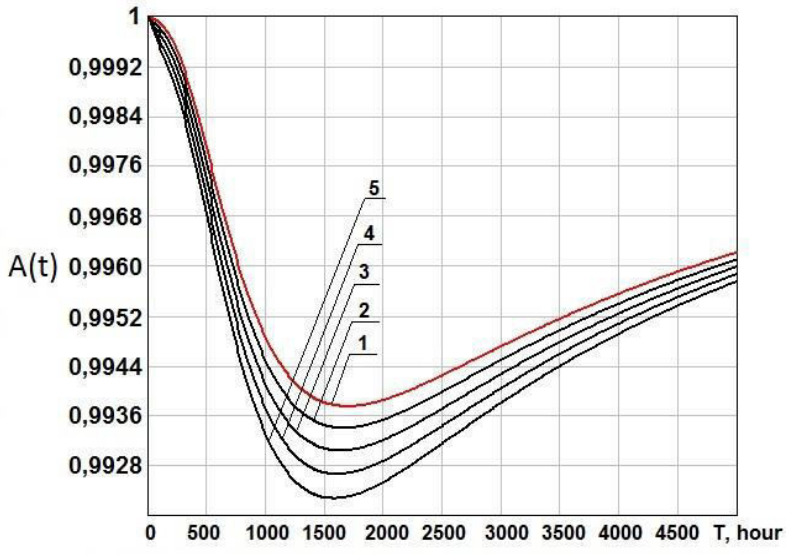


Fig. 4. Dependencies of availability function of the FTCS on values of the software update durations (duration of the first software update for 10 to 50 hours; the duration of the second firmware update - 200 hours).

The second experiment is conducted for the condition where the duration of the first SW update is significantly longer then the duration of the second SW update. The duration of the first update is 200 hours and the duration of the second update is given within 10 - 50 hours. The experiment is conducted with the following parameters FTCS: $\lambda_{hw}=1\cdot10^{-5}$ hour^{-1}; $\lambda_{sw11}=2\cdot10^{-3}$ hour $^{-1}$, $\lambda_{sw12}=1\cdot10^{-3}$ hour^{-1}; $\lambda_{swerror}=1\cdot10^{-2}$ hour^{-1}; T_{rest}=6 min; T_{switch}; T_{rep}=200 hour; T_{up1}=200 hour; (*line 1* - T_{up2}=10 hour; *line2* - T_{up2}=20 hour; *line3* - T_{up2}=30 hour; *line4* - T_{up2}=40 hour; *line5* - T_{up2}=50 hour).

The following results are produced by the proposed experiments:

1) If the duration of the first software update is significantly shorter than the duration of the second update, a decrease of the availability function of the readiness of the operational interval to 1700 hours is observed. If the duration of the first software update is significantly longer the decrease of the availability function, are readiness of the operational interval to 800 hours is observed.

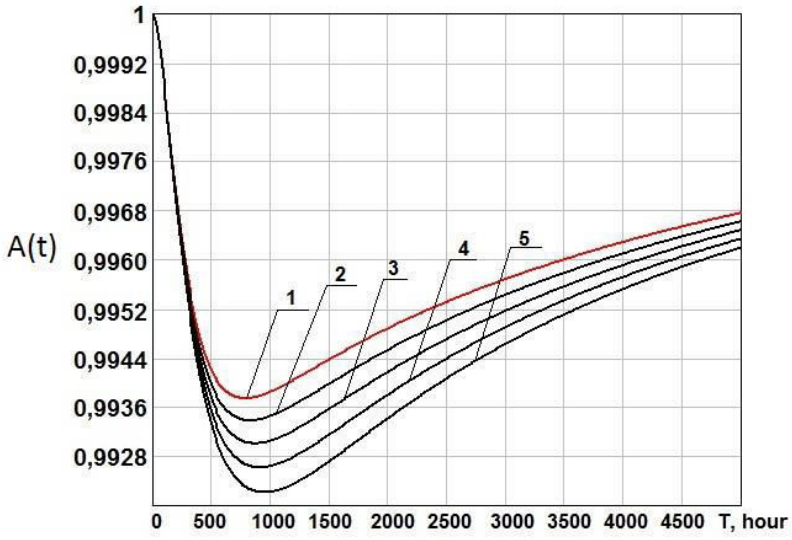


Fig. 5. Dependencies of availability function of the FTCS on the values of software update duration (duration of the first software update - 200 hours; the duration of the second firmware update from 10 to 50 hours).

2) The minimal decrease level of the availability function of the readiness of FTCS in the first and the second experiments is the same. This is explained by the values chosen for the duration for the first and the second software updates.

3) With the assistance of the proposed model it is possible to choose the duration of software updates that helps to ensure a minimum allowed level of the decrease of the availability function of the FTCS.

6 Conclusion

This research presents a model of FTCS with version-structural redundancy, with software updates and restart for automated development of Markovian chains using a unique technology and specific tool ASNA.

The presented model can be easily adapted to different configurations of FTCS, which envisages the use of the majority voting reservation in the hardware part and as a consequence in the majority of software versions from different developers. This model can be adopted for an unlimited number of software updates.

Future research has the potential to supplement this model with further factors:

- Erlang distribution for durations of software updates [15];
- unsuccessful restarting; unreliable commutation of elements and so on.

References

1. Mudry, P.A., Vannel, F., Tempesti, G., Mange, D. A Reconfigurable Hardware Platform for Prototyping Cellular Architectures. In: International Parallel and Distributed Processing Symposium. IEEE International, pp. 96–103 (2007)
2. Viktorov, O. Reconfigurable Multiprocessor System Reliability Estimation. Asian Jounal of Information Technology 6 (9),958–960 (2007)
3. Rajesh, S., Vinoth Kumar C., Srivatsan, R., Harini, S., Shanthi, A. Fault Tolerance in Multicore Processors With Reconfigurable Hardware Unit. In: 15th International conference on high performance computing. Bangalore, INDIA, pp. 166–171 (2008)
4. Amerijckx, C., Legat, J.-D. A Low-Power Multiprocessor Architecture For Embedded Reconfigurable Systems. In: Power and Timing Modeling, Optimization and Simulation, International Workshop, pp. 83–93 (2008)
5. Changyun Zhu, Gu, Z., Dick, R., Shang, L. Reliable Multiprocessor System-On-Chip Synthesis. In: International Conference Hardware/Software Codesign and System Synthesis, pp. 239–244 (2007)
6. Kim P. Gostelow. The Design of a Fault-Tolerant, Realtime, Multi-Core Computer System. In: In Aerospace Conference, IEEE, pp. 1–8 (2011)
7. Lyu M.R. (ed.), Software Fault Tolerance, New York: John Wiley & Sons (1995)
8. Korotun, T.M. Models and Methods for Testing Software Systems. Programming problems 2, 76–84 (2007) (In Russian)
9. Volochii, B.: Technology of Modeling the Information Systems. Publishing NU "Lviv Polytechnic" (2004) (In Ukranian)
10. Lei Xiong, Qingping Tan, Jianjun Xu. Effects of Soft Error to System Reliability. In: Workshops of International Conference on Advanced Information Networking and Applications. pp. 204–209 (2011)
11. Ponochonvyi, J.L., Odarushchenko, E.B. The Reliability Modeling Non-Redundant Information and Control Systems with Software Updated. Radioelectronic and computer systems 4(8), 93–97 (2004) (In Russian)
12. Kharchenko, V., Sklyar, V., Volkoviy, A.: Development and Verification of Dependable Multi-Version Systems on the Basic of IP-Cores. Proc. Int. Conf. Dependability of Computer Systems (2008)
13. Kharchenko V., Ponochovny Y., Boyarchuk A., Ermolayev V.: Availability Assessment of Information and Control Systems with Online Software Update and Verification. In: Information and Communication Technologies in Education, Research, and Industrial Applications Communications in Computer and Information Science, Vol. 469, Springer International Publishing Switzerland, pp. 300-324 (2014)
14. Moranda, P. B. An Error Detection Model for Application During Software Development. IEEE Trans. Reliability 4, 309–312 (1981)
15. Bobalo, J., Volochiy. B., Lozynskyi, O., Mandzii, B., Ozirkovskyi, L., Fedasuk, D., Shcherbovskyh, S., Yakovyna, V. Mathematical Models and Methods for Reliability Analysis of Electronic, Electrical and Software Systems, Lviv Polytechnic Press (2013)

Features of Hidden Fault Detection in Pipeline Digital Components of Safety-Related Systems

Alex Drozd, Miroslav Drozd, Viktor Antonyuk

Institute of Computer Systems, Odessa National Polytechnic University,
ave Shevchenko 1, 65044 Odessa, Ukraine
Drozd@ukr.net, miroslav_dr@mail.ru, melmoth@te.net.ua

Abstract. Paper is devoted to a problem of the hidden faults, which are appropriate for safety-related instrumentation and control systems aimed at ensuring the safety of high-risk objects. Such systems are designed for operation in two modes: normal and emergency. The problem consists in accumulation (during a normal mode) of the hidden faults impairing the functionality of the digital components and system in an emergency mode. A model of activated path that determines the input data for simulation of pipelined digital component is offered. Simulation is executed to assess observability of the circuit points and to detect the potentially hazardous points, which are carriers of considered faults. The method of identifying potentially hazardous points in circuits with the LUT-oriented architecture in FPGA projects of digital components is proposed.

Keywords. safety-related instrumentation and control system, pipeline digital component, hidden faults, controllability, observability

Key Terms. HighPerformanceComputing, ConcurrentComputation, Model, Method, Simulation

1 Introduction

The high-risk objects presented in energy, on transport, in space and defense branches have become an essential part of human environment. These include the power grid and power plants, aircraft and ground systems of ensuring flights, various kinds of weapons. Development and exploitation of these objects is impossible without wide use of information technologies which act as a counterbalancing factor of the complexity growing quantitatively and qualitatively, power and danger of critical applications [1].

The safety-related instrumentation and control systems (I&CS) which are the development of computer systems with the diversification of an operation mode by its division into normal and emergency are designed for servicing of high-risk objects. Great demands for a complex of attributes which are regulated by the international standards are made of I&CS. Requirements for ensuring functional safety of I&CS based on the construction of fault-tolerant components are distinguished from the most important [2].

The technologies of design of the fault-tolerant digital devices including use of the correcting codes, majority structures, different types of element reservation and system reconfiguration and also the multi-version solutions for prevention of faults caused by the common reason are traditionally applied to the digital components (DC) [3].

However, the fault tolerance of I&CS and its components cannot be provided in a separation from a solution of a problem of the hidden faults which can accumulate during a long normal mode in DC circuits owing to their low checkability [4].

In practice the problem of hidden faults is solved by improving the checkability of DC using periodic checking [5]. It is performed in testing by imitation of emergency mode with shutdown of emergency protection. On-line testing is used in periodic checking at manual regulation of the input data with the approximation to the conditions of the emergency mode, keeping within the normal mode. This solution has often led to emergency consequences of unauthorized inclusion of imitation in emergency mode by the person or fault [6]. Manual regulation and shutdown of emergency protection preceded Chernobyl catastrophe.

Thus, the problem of hidden faults is better known for emergencies that arise due to actions aimed at its solution. Faults remaining hidden have not led to any accidents. At the same time, the history of fight against them shows mistrust to fault tolerance of I&CS and its components.

For verification of I&CS, its components and solutions used for their development and testing apply a number of methods and technologies, including [7-9]:

- Expanded Functional Testing to study the behavior of I&CS on the occurrence of rare events, for example, multiple failure;
- Event Tree Analysis (ETA) and Fault Tree Analysis (FTA), considering the sequence of events and fault developing in ICS;
- Failure Modes, Effects and Criticality Analysis (FMECA) of components on their criticality for safety of I&CS (it is aimed at determining the need for special conditions of design and operation);
- Fault Insertion Testing (FIT) to evaluate the methods and means of testing and the consequences caused by the fault.

These and other measures stipulated by international standards in the existing I&CS does not directly put and do not solve the problem of hidden faults. In practice, these measures do not solve the full problem of functional safety systems and facilities management. Proof of that are numerous accidents in recent years in power networks and power plants, train wreck and the crash, failed launches of spacecraft.

The concept of checkability was formed in testing as testability for estimating the complexity of test generation and further testable design of the digital devices aimed at detecting faults in pauses of an operating mode. The assessment is carried out for points of the digital circuit by calculation of controllability, observability and checkability as their product [10]. In on-line testing observability coincides with checkability, and controllability is their upper bound. The checkability of DC becomes structurally functional showing the dependence not only on the structure of the circuit, but also the characteristics of the input data [11]. On its basis the methods for improving checkability in normal mode of I&CS by approach to the upper bound and its raising are developed [12].

In I&CS the checkability of DC is diversified, becoming different for normal and emergency mode. A model of dual-mode structural-functional checkability is offered in [13]. On its basis controllability and observability of a point of the circuit is diversified and the method of detection in the DC circuit of the potentially hazardous points (PHP) in which during a normal mode there can be a hidden fault reducing fault tolerance of the DC upon transition of I&CS to emergency mode is offered. It is proposed to identify PHP on their observability which is defined for normal and emergency mode by simulating the operation of simultaneous DC in the ranges of input data of these modes. The method allows estimating the circuit of the DC in probability of reducing the fault tolerance of the DC upon transition to emergency mode, whereas the percentage of possible hidden faults of certain type [14].

The offered paper is aimed at development of models and methods for determining the observability to identify PHP for pipeline DC used in I&CS. Section 2 discusses a model of activated path that determines the DC input data required to assess the observability of the circuit point. A method of pipeline DC simulation for an assessment of the observability of the circuit points at the input data processed in a mode of I&CS and detection of the PHP is proposed in section 3. A method of detecting internal PHP of the DC circuit with LUT-oriented architecture in FPGA projects is offered in section 4.

2 The Model of the Activated Path of Pipeline DC Circuit

Typically, DC for I&CS are built under construction the pipeline. The sections of pipeline are simultaneous units that perform one or more arithmetic and logical operations with numbers represented in parallel codes. In each clock cycle the input words made up of operands of the operations, including the processing numbers and control bits come at the inputs of pipeline DC.

For stuck-at faults, belonging of a circuit point to a set of PHP is completely determined by the values of its observability in normal and emergency mode, using the formulas [14], or as it is shown in Table 1.

Table 1. Conditions of circuit point belonging to a set of PHP

Mode	Emergency		
Normal	$O_E = 1$	$O_E = 2$	$O_E = 3$
$O_N = 2$	‘1’	–	‘1’
$O_N = 1$	–	‘0’	‘0’
$O_N = 0$	‘1’	‘0’	‘1’, ‘0’

Rows and columns of Table 1 contain the values of observability O_N and O_E for normal and emergency modes of I&CS, respectively. At their intersection the types of stuck-at faults: ‘0’ or ‘1’, which include the point of the circuit to the set of PHP are shown. Observability O_N or O_E of a point takes on the values 0, 1, 2 or 3 if this point is not observable, observed in ‘1’ value, ‘0’ value or both values ‘0’ and ‘1’, respectively.

The point is observable or not observable in the presence or absence of the path activated from this point to a check point of the circuit. The path is activated if an erroneous value accepted by a signal in the point passes this path at the input data of the considered I&CS mode.

Table 1 shows a necessary condition for the point belonging to set of PHP. Within this condition the PHP are identified on condition $O_N + O_E = 3$ according to which the quantity N_{PHP} of PHP in the circuit of the DC can be estimated. Two types of stuck-at faults can arise in PHP when performing a condition $(O_N = 0) \vee (O_E = 3)$. Their quantity N_{DHF} allows estimating probability of decrease in fault tolerance of the DC upon transition to emergency mode [14] by the formula

$$P_{FTR} = (N_{PHP} + N_{DHF}) / (2N_{DCC}), \quad (1)$$

where N_{DCC} is the total number of points of the DC circuit.

To identify PHP of the circuit it is necessary to properly assess their observability O_N and O_E, whereas the DC input data, typical for normal and emergency mode of I&CS.

We propose the following model of the activated path AP of DC circuit: $AP(U_1, \ldots, U_I, \ldots, U_K)$, where U_I, ..., U_K are descriptions of data processing units on the pipeline sections that make up the path AP; K is the number of units U_I, $I = \overline{1, K}$. Each unit U_I is represented by a model $U_I (F_I, Z_{I1}, \ldots, Z_{IJ}, \ldots, Z_{IM_I})$, where F_I is the operation performed by the unit U_I; $Z_{I1}, \ldots, Z_{IM_I}$ – descriptions of inputs of the unit U_I; M_I is the number of inputs of the unit U_I. Each input Z_{IJ} is characterized by the distance D_{IJ} between own and current word and also by values calculated at the input words of the DC circuit. The distance is measured in clock cycles according to the formula

$$D_{IJ} = | Y_I - Y_{IJ} |, \quad (2)$$

where Y_I is the number of the current word input to the unit U_I by input Z_I belonging to the way AP; Y_{IJ} is the number of own word at the input Z_{IJ}. Input Z_I is one of the inputs of Z_{IJ} for which $D_{IJ} = 0$.

In case $\forall(I, J)$, $D_{IJ} = 0$ unit U_I is simulated at the inputs Z_{IJ}, each of which receives the value calculated for the current word. Model of activated path AP is simplified so that the evaluation of the observability can accumulate with consecutive simulation of path AP at separate words of I&CS mode.

In case $\exists(I, J)$, $D_{IJ} > 0$ the unit U_I is simulated at the inputs Z_{IJ}, which takes the value calculated for the current word, and the word input to the DC input on D_{IJ} clock cycles sooner or later. Generally this word can be any input word of the considered mode. Therefore, the unit U_I should be simulated for each of the input words on the values of input Z_{IJ} which are accepted by it on all words of the considered mode.

For inertial processes of change of the input data processed in DC, the model of path AP can be refined taking into account the maximum possible step Δ of changing the input words or the operands making them. Let $\Delta \geq 1$. Then the unit U_I should be simulated for each input word with number G on the values of input Z_{IJ}, which are

accepted in the range of words with numbers $G \pm \Delta \cdot D_{IJ}$,, where $G > \Delta \cdot D_{IJ}$. For $G \leq \Delta \cdot D_{IJ}$ the initial set value of input D_{IJ} are used. In case of $\Delta < 1$, when the value of the input word is changed no more than once per $1 / \Delta$ clock cycles, the number range is rounded to the value of $G \pm] \Delta \cdot D_{IJ} [$.

Building a model of the path AP is performed by analyzing the structure of the pipeline taking into account quantity of the sections preceding the data processing unit U_I on its inputs Z_{IJ}. The quantity of these sections determines amount of clock cycles required for data delivery from the inputs of the DC circuit to the unit U_I. Let H_{IJ} is the quantity of sections (clock cycles) preceding the input Z_{IJ}, and $H_{MAX} = \mathrm{MAX}(H_{IJ})$. Then, the equality $H_{IJ} + Y_{IJ} = H_{MAX} + 1$ is carried out for unit U_I in clock cycle H_{MAX}. This determines number of own word

$$Y_{IJ} = H_{MAX} + 1 - H_{IJ}, \tag{3}$$

including current word Y_I for input Z_I.

Substituting (3) in (2) determines the values of the distance $D_{IJ} = | H_I - H_{IJ} |$ in the model of path AP by the structure of the pipeline DC.

3 Simulation of the Circuit of a Pipeline DC

The observability of the circuit point is calculated in the course of pipeline DC simulation at the input data, which are determined taking into account the model of path AP.

Simulation of DC is performed according to the following method. Examination of all input words of the considered mode and examination of all points of the circuit on the pipeline course will be organized. The value of the examined point is calculated at a given input word and is complemented by an inverse value. Values of all following points of the circuit are calculated for two values of the examined point before reaching a check point or a point where results of calculations coincide. If the results in a check point are inverted, all the points belonging to the path, refer to the 0-observable or 1-observable depending on the values accepted by them. If at the previous input words the point was identified as observed with opposite value, this point refers to observable and is not considered at the following input words. Point values with $D_{IJ} > 0$ are calculated on an extended set of input words. The simulation is carried out taking into account all their values.

It should be noted that incomplete simulation of the DC circuit in case of a restriction of the input data typical for considered mode leads to an underestimation of the observability of points. This underestimation admitted for normal and emergency mode conducts to false detection of PHP and their skipping, respectively.

Considering that controllability of the circuit points is estimated directly by their values, i.e. it is much simpler than observability, and taking into account that controllability is the upper bound of observability, the following method of PHP identification is offered:

- Comparison in sizes $| R_N |$ and $| R_E |$ of ranges R_N and R_E for the input data of normal and emergency modes.

- In case $|R_N| > |R_E|$ simulation of the DC is running at the input words of emergency mode for determining the sets E_{O-0}, E_{O-1}, E_{O-2} and E_{O-3} of points with observability of O_E: 0, 1, 2 and 3. The points of the set E_{O-0} are excluded from consideration as they cannot be PHP. For the other points, the simulation of DC is running at the input words of normal mode for determining the sets N_{O-0}, N_{O-1}, N_{O-2}, N_{O-3} of points with observability of O_N: 0, 1, 2, 3 and identification of PHP, according to Table 1.
- In case $|R_N| \leq |R_E|$ simulation of the DC is running at the input words of normal mode for determining the sets N_{O-0}, N_{O-1}, N_{O-2} and N_{O-3} of points with observability of O_N: 0, 1, 2 and 3. The points of the set N_{O-3} are excluded from consideration. The simulation of the DC is running at the input words of emergency mode to determine the sets E_{C-1} and E_{C-2} of points with controllability of C_E: 1 and 2. The points of the sets $N_{O-1} \cap E_{C-1}$ and $N_{O-2} \cap E_{C-2}$ are excluded from consideration. For the other points, simulation of the DC is running at the input words of emergency mode for determining the sets E_{O-0}, E_{O-1}, E_{O-2}, E_{O-3} of points with observability of O_E: 0, 1, 2, 3 and identification of PHP, according to Table 1.

4 Identification of Internal PHP in Circuits with the LUT-Oriented Architecture

Modern I&CS are designed with the use of pipeline DC constructed on FPGA with the LUT-oriented architecture. The feature of the circuits of such DC consists in a table specifying logical functions in memory of LUT (Look-Up Table). The result of function is read out (with use of the multiplexer) from memory of LUT with the address which code is formed from arguments of function [15].

The bits of the LUT memory are considered as internal points of the DC circuit. Stuck-at faults of an internal point can be caused by defect of memory bit or defect of the multiplexer. The internal point of the circuit is controllable if the appropriate bit is selected at the input data of the considered mode, and is uncontrollable otherwise. The internal point is observable if the point of the LUT output is observable at a choice of the appropriate bit of memory.

The set of all controllable internal points of the circuit can be calculated for each mode by simulation of DC at all input words of this mode including additional words for points with $D_{IJ} > 0$. All internal points addressed in memory of LUT belong to the controllable.

Internal PHP of the circuit can be identified according to the following method: For each LUT two sets N_C and E_C of internal points which are controllable respectively in normal and emergency mode are determined. The set of $C_{EN} = E_C \setminus N_C$ of the internal points addressed in emergency mode and not used in the normal one is calculated. Internal points of a set of C_{EN} are checked for observability in emergency mode. If they are observable, refer to the set of PHP.

For example, the task of identifying internal PHP in the circuit of DC that computes the function $F(X) = 1$ for $X \bmod 3 = 0$ and $F(X) = 0$ for other values of X, where $X = \{x_5, x_4, x_3, x_2, x_1\}$, $X = 0 \div 31$. The circuit is evaluated for stuck-at faults for the given ranges R_N and R_E of input data.

The solution is considered for the three DC functioning in I&CS in different conditions:

a) $R_N = 0 \div 23$ and $R_E = 24 \div 31$;
b) $R_N = 0 \div 15$ and $R_E = 16 \div 31$;
c) $R_N = 8 \div 23$ and $R_E = 0 \div 7, 24 \div 31$.

A description of the function F and ranges R_N and R_E are shown in Table 2.

Table 2. Description of the function $F(x_5, x_4, x_3, x_2, x_1)$ and ranges R_N and R_E

X for x_5		Variables				F for x_5		Ranges R_N and R_E					
1	0	x_4	x_3	x_2	x_1	1	0	a		b		c	
16	0	0	0	0	0	1	0	0	16	0	16	0	16
17	1	0	0	0	1	0	0						
18	2	0	0	1	0	0	1						
19	3	0	0	1	1	1	0						
20	4	0	1	0	0	0	0		R_N			R_E	R_N
21	5	0	1	0	1	0	1						
22	6	0	1	1	0	1	0						
23	7	0	1	1	1	0	0		23			7	23
24	8	1	0	0	0	0	1	R_N	24	R_N	R_E	8	24
25	9	1	0	0	1	1	0						
26	10	1	0	1	0	0	0						
27	11	1	0	1	1	0	1		R_E			R_N	R_E
28	12	1	1	0	0	1	0						
29	13	1	1	0	1	0	0						
30	14	1	1	1	0	0	1						
31	15	1	1	1	1	1	0	15	31	15	31	15	31

The number of input words and the function values are shown in pairs of columns separately for values $x_5 = 0$ and $x_5 = 1$. The ranges R_E of emergency mode selected dark color.

The circuit of DC designed on FPGA ALTERA Quartus II [16] is shown in Fig. 1.

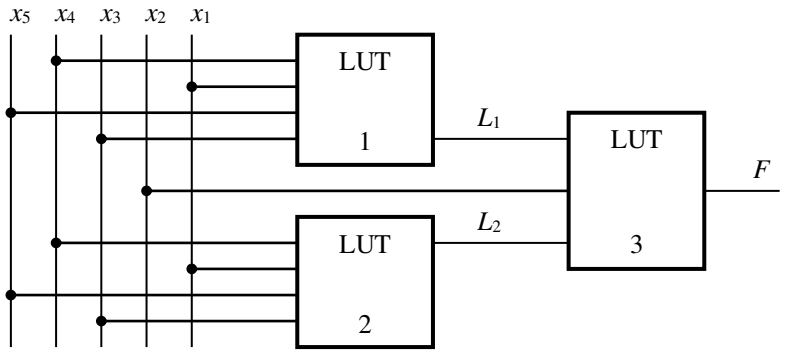


Fig. 1. The circuit of DC

The circuit consists of three LUT: LUT 1, LUT 2 and LUT 3, which implement the functions $L_1(x_3, x_5, x_1, x_4)$, $L_2(x_3, x_5, x_1, x_4)$ and $F(L_2, x_2, L_1)$ described by codes $6BBD_{16}$, $97E9_{16}$ and $0AA0_{16}$, respectively.

Values of the variables x_3, x_5, x_1, x_4, arriving at the inputs of the LUT 1 and LUT 2 in emergency mode, and also calculated values of the functions L_1, L_2, F and F_{L1}, F_{L2} are shown in Table 3.

Table 3. Description of input variables and functions for LUT 1 and LUT 2

№	Variables				L_1	L_2	F for x_2		F_{L1} for x_2		F_{L2} for x_2		X for x_2	
	x_3	x_5	x_1	x_4			0	1	0	1	0	1	0	1
0	0	0	0	0	0	0	1	0	**0**	0	0	0	0	2
2	0	0	1	0	1	0	0	1	0	**0**	1	0	1	3
4	0	1	0	0	1	0	0	1	0	**0**	0	**0**	16	18
5	0	1	0	1	0	0	1	0	**0**	0	**0**	**1**	24	26
6	0	1	1	0	0	1	0	0	**1**	0	0	0	17	19
7	0	1	1	1	1	0	1	0	**0**	0	**0**	0	25	27
8	1	0	0	0	1	0	0	1	0	**0**	**1**	**0**	4	6
10	1	0	1	0	0	1	0	0	**1**	0	0	0	5	7
12	1	1	0	0	0	1	0	0	**1**	0	0	0	20	22
13	1	1	0	1	1	0	0	1	0	**0**	**1**	**0**	28	30
14	1	1	1	0	1	1	1	0	**0**	0	**0**	**1**	21	23
15	1	1	1	1	0	1	0	0	**1**	0	0	0	29	31

The functions F_{L1} and F_{L2} take the values of the function F for inverse value according to L_1 and L_2 owing to the distorted value of internal point in the LUT memory. Besides, the first column contains the numbers of sets of the input variables equal to their decimal equivalent. They are also numbers of internal points disposed in the LUT memory. The values of X are specified in two last columns for two values of variable $x_2 = 0$ and $x_2 = 1$. Values of functions F_{L1}, F_{L2} selected in Table 3 by dark color are inversed to values of function F that defines the internal points of LUT corresponding to them as observable.

According to Table 3 internal points which are observable in emergency mode for the considered ranges of R_N and R_E compose the following sets:

a) 5, 7, 13, 15 to LUT 1 and 5, 7, 13 to LUT 2;
b) 4 – 7, 12 – 15 to LUT 1 and 4, 5, 7, 13, 14 to LUT 2;
c) 0, 2, 5, 7, 8, 10, 13, 15 to LUT 1 and 0, 2, 5, 7, 8, 13 to LUT 2.

Listed internal points are not used in the normal mode, i.e. are unobservable. Therefore they belong to the set of PHP. In LUT 3 the same 6 internal points 1, 3 – 7 in all modes are used that excludes them from a set of PHP as $C_{EN} = \varnothing$ and the output of LUT 3 is an observable point. The quantity of PHP in DC for the considered three cases makes respectively 7, 13 and 14 at total of internal points $N_{DCC} = 40$. The probability of decrease in fault tolerance of the DC upon transition to emergency mode calculated by the formula (1) accepts values of 17.5%, 32.5% and 35% for cases of a, b and c, respectively.

5 Conclusions

Two modes of operation characteristic for I&CS generate the problem of hidden faults which can accumulate in normal mode and reduce fault tolerance of the DC in the most responsible emergency mode. In single-mode systems such problem isn't present as a hidden fault is never shown and if a fault was shown, it isn't hidden.

The success in the solution of a task of PHP identification where the problem of the hidden faults occurs is determined by opportunities of an assessment in observability of circuit points in each mode of I&CS. The correctness of this assessment is provided with completeness of input data considered for each mode of I&CS. The underestimated assessment leads to false detection of PHP or their skipping. The offered model of the activated path specifies a set of input data for determination of observability of points in circuits of pipeline DC. The method of simulation of pipeline DC and method of PHP identification follows from this model.

The design of modern I&CS on FPGA defines an additional task of identifying internal PHP typical for circuits with LUT-oriented architecture. The offered method of solving this task analyzes sets of internal points used in normal and emergency mode.

It should be noted that these sets determine the effectiveness of manual regulation of the input data in the normal mode. This procedure is used in practice for solving the problem of the hidden faults. Procedure reduces the effectiveness with different sets and becomes completely meaningless in case of disjoint sets.

References

1. Bakhmach, E., Kharchenko, V., Siora, A., Sklyar, V., Tokarev, V.: Design and Qualification of I&C Systems on the Basis of FPGA Technologies. In: 7th International Topical Meeting on Nuclear Plant Instrumentation, Control, and Human-Machine Interface Technologies (NPIC&HMIT 2010), pp. 916–924. Las Vegas, Nevada (2010)
2. IEC 61508-1:2010. Functional Safety of Electrical / Electronic / Programmable Electronic Safety Related Systems – Part 1: General requirements. Geneva: International Electrotechnical Commission (2010)
3. Sklyar, V.V., Kharchenko, V.S.: Fault-Tolerant Computer-Aided Control Systems with Multiversion-Threshold Adaptation: Adaptation Methods, Reliability Estimation, and Choice of an Architecture. Automation and Remote Control 63(6), 991–1003 (2002)
4. Drozd, M., Drozd, A.: Safety-Related Instrumentation and Control Systems and a Problem of the Hidden Faults. In: 10th International Conference on Digital Technologies, pp. 137–140. Zhilina, Slovak Republic (2014)
5. Kharchenko, V.S., Sklyar, V.V. (eds): FPGA-based NPP I&C Systems: Development and Safety Assessment: RPC Radiy, National Aerospace University "KhAI", SSTC on Nuclear and Radiation Safety, Kharkiv, Ukraine (2008)
6. Gillis, D. The Apocalypses that Might Have Been, http://www.popmech.ru/go.php?url=http%3A%2F%2Fwww.damninteresting.com%2F%3Fp%3D913
7. Garcia, P.A., Schirru, R., Frutuoso P.F., Melo, E.: A Fuzzy Data Envelopment Analysis Approach for FMECA. Progress in Nuclear Energy 46, 359–373 (2005)
8. Andrashov, A., Kharchenko, V., Sklyar, V., Reva, L., Dovgopolyi, V., Golovir, V.: Verification of FPGA Electronic Designs for Nuclear Reactor Trip Systems: Test- and Invariant-Based Methods. In: IEEE East-West Design & Test Symposium (EWDTS'10), pp. 92–97. St. Petersburg, Russia (2010)

9. Kharchenko, V.S. (ed): Safety of Critical Infrastructures: Mathematical and Engineering Methods of Analysis and Ensuring: Ministry of Education and Science, National Aerospace University "KhAI", Kharkiv, Ukraine (2013)
10. IEEE 1149-1:2001, IEEE Standard Test Access Port and Boundary-Scan Architecture. IEEE Computer Society (2001)
11. Drozd, A., Kharchenko, V., Antoshchuk, S., Sulima, J., Drozd, M.: Checkability of the Digital Components in Safety-Critical Systems: Problems and Solutions. In: IEEE East-West Design & Test Symposium, pp. 411–416. Sevastopol, Ukraine (2011)
12. Drozd, A., Kharchenko, V., Antoshchuk, S., Drozd, J., Lobachev, M., Sulima, J.: The Use of Natural Resources for Increasing a Checkability of the Digital Components in Safety-Critical Systems. In: IEEE East-West Design & Test Symposium, pp. 327–332. Kharkiv, Ukraine (2012)
13. Drozd, A., Kharchenko, V., Antoshchuk, S., Drozd, M.: Checkability of Safety-Critical I&C System Components in Normal and Emergency Modes. Journal of Information, Control and Management Systems 10(1), 33–40 (2012)
14. Drozd, A., Drozd, M.: A New Approach to Solving a Problem of the Hidden Faults in Safety-Related Systems Journal of Information, Control and Management Systems 12(2), 125–132 (2014)
15. Cyclone FPGA Family Data Sheet. Altera Corporation (2003), http://www.altera.com
16. Netlist Optimizations and Physical Synthesis. Qii52007-2.0. Quartus II Handbook. Vol. 2. Altera Corporation (2004)

The Control Technology of Integrity and Legitimacy of LUT-Oriented Information Object Usage by Self-Recovering Digital Watermark

Kostiantyn Zashcholkin and Olena Ivanova

Odessa National Polytechnic University, 1, Shevchenko Avenue, Odessa, Ukraine
(const-z@te.net.ua, en.ivanova.ua@gmail.com)

Abstract. The paper proposes the technology of control of integrity and legitimacy of information object usage with the Look Up Table-oriented (LUT-oriented) architecture. The technology is based on embedding the self-recovery digital watermark into information objects of such kind. The technology is the composition of approaches to forming the self-recovery digital watermarks in the passive multimedia containers, and approaches to embedding the extra information into LUT-containers. The process of embedding the extra information occurs with the help of classification of container elements and their purposeful modification within the set of the formed classes. The procedure of immediate embedding the data at the level of LUT-container elementary parts includes the value inversion of current processed LUT unit and propagation of the inversion around all the inputs of LUT units connected to the current unit output. The description of practical realization of the proposed technology is represented.

Keywords: digital watermarks, control of information object integrity, control of information object usage, information security, cybersecurity, IT systems safety, steganography, LUT-oriented architecture, FPGA.

Key terms: Information Technology, Data, Object, Approach, Method.

1 Introduction: Topicality of the Problem and Aim of the Paper

The technologies of digital watermark (DWM) usage are one of the most effective approaches to the comprehensive information security provision [1]. DWMs are mainly used for:

1. the control of information object integrity;
2. the control of information object usage: information object copyright provision, digital content authentication, the tracking of digital content move (including the tasks of source retrieval of information leak).

DWM technologies are based on steganographical technique, with the help of which the fact of DWM presence in an information object (DWM container) is

hidden. At the same time DWM can be read in the container if someone has a stego-key possessing the set of access rules to DWM elements [2].

The *control of information object integrity* by DWM is based on embedding some control data unit allowing to analyze the object integrity in an information object. Hash sum calculated with the help of the definite hash-function is most frequently used as such kind of data unit [3]. However such kind of control unit embedding results in information object change and consequently violates its integrity by itself. Under these conditions the so-called *self-recovering DWMs* possessing the ability to recover the initial object value (a value, which it had before DWM embedding) in the process of DWM reading from the object are used [4]. In order to control the integrity the DWM extraction from information object, hash sum calculation and this hash sum with DWM contents comparison is performed. The analysis of integrity control guarantees the impossibility of the information object substitution or corruption. Finally this provides the *safety of functioning the information system* processing the given information object.

The control of information object usage legitimacy by DWM is based on possibility of a copyright owner to embed DWM having the copyright owner identifying data into the object. So it is only a copyright owner who possesses a stego-key can check the DWM presence and its contents in the information object [5]. DWM technologies are actively used in the structure of Digital Rights Management systems (DRM – systems), but the field of their usage is commonly limited with the multimedia content protection: graphical, audio and video files [6], [7]. In performing the tasks of control of information object usage legitimacy by DWM the recovery of initial object value after DWM extraction is necessary.

In the present paper the author proposes the information technology of self-recovery DWM usage in active hardware containers based on LUT-oriented architecture (further LUT-containers). We can refer, for example, Field Programmable Gate Array microchips (FPGA microchips) [8], which at present are often used as an element base for computer and control system design, to such kinds of containers. The main element of such containers is LUT units, which are the data structure used to replace the calculation with prepared data search operation [9]. LUT units in FPGA are normally represented in the form of RAM. In this case LUT unit inputs are the address inputs of RAM. If the number of inputs equals n a LUT unit stores 2^n bits information and is capable of calculating the value of 1 n-argument Boolean function.

Self-recovering DWMs have been developed and used in the field of control of integrity and legitimacy of usage for passive multimedia containers, e.g. graphical, audio and video files. However it is necessary to control not only the information objects of such kind by DWMs. But the task of self-recovering DWM usage in active containers performing some calculating or controlling function is not realized at the moment. In the articles [10], [11] the techniques of DWM embedding in active LUT-containers are proposed. But the techniques described do not possess the possibility to recover the container original. So *the aim of the given paper* is to describe the possibility of container original recovery after DWM extraction by developing the technique of forming the DWMs in LUT-containers.

2 The Information Technology of Embedding the Self-Recovering DWM in LUT-Container

The information technology considered in the paper is a formal process of usage of the proposed and already existing techniques as well as the means of information processing, which provides DWM forming in LUT-container space.

The proposed information technology uses a set of *Fridrich-Goljan-Du* method (further *Fridrich* method) approaches [12], [13] in the part, which is intended for container recovery, and technique described in [10], [11] in the part intended for DWM embedding in LUT-containers. Besides the represented technology uses such peculiarities of LUT-containers as activeness, accurate data presentation, non-autonomousness of their elementary parts [10].

The methods proposed in [10], [11] are based on the change of codes of a pair of successively integrated LUT units, which does not alter this pair functioning. Assume there is a pair of LUT units and the output of the first one is connected directly or through a trigger to the address input of the second one. The inversion of all the bits of the first unit code and a definite rearrangement of the bits of the second unit code does not change the functioning of the given pair of units. It is the consequence of equivalence of Boolean functions realized by the pair of units before and after values inversion. The rearrangement of bits in the second unit code of the pair is produced according to the rules depending upon the binary weight of its address input, which the first unit output is connected to. The inversion of bits of the first unit code of the pair gives the possibility to achieve the required value in a definite bit of a unite code. This peculiarity is exactly used by the methods considered in [10], [11] for DWM embedding in a LUT-container. However these methods are not able to recover the container original.

Fridrich method [12], [13] is based on the group processing – disjoint subsets of elementary parts of a container with the help of two functions: Flipping-function F and function of discrimination f.

Function of discrimination is used in Fridrich method on order to classify the groups of elementary container parts by including them to the classes of regular, singular and unused groups. To determine the function value in each of the container groups we are to calculate the sum of pairwise subtractions with overlapping for elementary container parts included in a group.

Flipping-function is used by Fridrich method to modify the groups according to some rule possessing the characteristic of involution. In the proposed information technology the LUT-container architecture peculiarities are taken into account and this fact differs it from Fridrich method in the following features:

- within the frames of the proposed technology in the course of DWM embedding the modification of elementary parts exactly (LUT units) is performed but not the groups of elementary parts of a container;
- within the frames of the proposed technology the calculation of Flipping-function $F(x)$ unlike Fridrich method contains the LUT unit value inversion according to the procedure mentioned in [10], [11]. The procedure is the invention of values of the current processed LUT unit and propagation the inversion around the inputs of all

the LUT units connecting to the current unit output. It is obvious that the Fridrich method requirement concerning the presence of involution characteristics is performed for the given type of Flipping-function, i.e. $F(F(\mathrm{LUT}_i)) = \mathrm{LUT}_i$ where LUT_i is any valid value of some LUT unit;

- unlike Fridrich method within the frames of the proposed technology a principle of container element classification without calculating any function of discrimination is determined. The element classification is organized as a determination of zero and one value proportion in LUT unit codes. It is performed according to the following principle: if LUT unit contains zero values more than one values it is classified as *a regular unit (R*-unit); if a unit contains more one values than zero values it is classified as *a singular unit (S*-unit); if the number of zero values of LUT unit equals the number of its one values the unit is classified as *an unused unit (U*-unit).

Certainly, the rules of moving the container from one class to another defined by Fridrich method are performed within the frame of the proposed approach as well:

$$\begin{aligned} F(\mathrm{LUT}_R) &= \mathrm{LUT}_S; \\ F(\mathrm{LUT}_S) &= \mathrm{LUT}_R; \\ F(\mathrm{LUT}_U) &= \mathrm{LUT}_U, \end{aligned} \tag{1}$$

where LUT_R, LUT_S, LUT_U – LUT units classified as *R*-unit, *S*-unit and *U*-unit, respectively; $F()$ – Flipping-function.

The succession of actions in the proposed information technology in DWM embedding in LUT-container is as follows.

Stage 1. The movement about LUT-container units in the order determined by embedding path is realized. During this action each of the LUT units refers to the classes *R*-unit, *S*-unit or *U*-unit on the basis of the above mentioned classification rules. According to the results of classification the binary *RS*-vector is formed in the following way: if the next LUT unit is classified as *R*-unit then zero value is fixed in *RS*-vector; if the next LUT unit is classified as *S*-unit then one value is fixed in *RS*-vector; *U*-unit values are not fixed in *RS*-vector.

Stage 2. The obtained *RS*-vector compression is performed with the help of some lossless compression algorithm. As a result of this the compressed vector RS_{com} is formed. It is obvious that RS_{com} vector length is less than *RS*-vector length. Difference of length of these vectors is denoted as ΔL.

Stage 3. DWM, which is a binary succession the length of which does not exceed ΔL is added to vector RS_{com} by concatenating. Vector *RS** obtained after concatenation is added with arbitrary binary values until it reaches the *RS* vector length. So:

$$RS^* = RS_{com} \,.\, DWM \,.\, Add, \tag{2}$$

where "." – operation of concatenation; *DWM* – watermark embedded into container; *Add* – optional addition of vector *RS** to *RS* vector length.

Stage 4. The movement about LUT units of container in the order determined by embedding path is performed. In the course of this movement *U*-units are ignored, and within *R*-units and *S*-units a mutual transition of one into another with the help of Flipping-function is made according to the table.

Table 1. The rules of LUT-unit move from one class to another

Class of the current LUT-unit LUT_i	Value of a bit, corresponding to the current LUT-unit LUT_i in vector *RS*	The current bit of vector *RS**	Action
R	0	0	—
R	0	1	Transformation LUT_i to *S*-unit
S	1	0	Transformation LUT_i to *R*-unit
S	1	1	—

The actions at the given stage leads to vector *RS** embedding in a LUT-container. Vector *RS** contains DWM and a compressed vector *RS* (vector RS_{com}) version in accordance with equation (2). Thus except for a DWM itself the information necessary for the original container value recovery is embedded in a container. The rules represented in table 1 describe the actions providing the vector *RS** embedding in a LUT-container. If the corresponding current bits of vectors *RS* and *RS** coincide then no modification for the current LUT unit is produced. In the case of mismatch of these bits the current LUT unit transition in the class corresponding to the current bit value of vector *RS** is performed.

Taking into account that the Flipping-function impact on LUT units occurs according to the unit inversion rules represented in [10], [11] LUT-container functioning does not change after DWM embedding.

3 Example of DWM Embedding According to the Proposed Technology

Let us consider the example of described stages of DWM embedding in container technology. In table 2 the hexadecimal values of codes of twenty LUT units ($i = 1…20$), which are in some embedding path of a container are shown in the lines "Unit code".

Table 2. Initial values of codes of LUT units lying in embedding path

i	1	2	3	4	5	6	7	8	9	10
Unit code	51F0	FF96	56EC	E8FE	2002	0008	8040	1000	0FF0	CCCD
n^1	7	12	9	11	2	1	2	1	8	9
Class of unit	*R*	*S*	*S*	*S*	*R*	*R*	*R*	*R*	*U*	*S*
RS-vector	0	1	1	1	0	0	0	0	—	1
i	11	12	13	14	15	16	17	18	19	20
Unit code	0220	88F8	F128	3F00	0001	9B94	AABA	A340	B0E1	EE00
n^1	2	7	7	6	1	8	9	5	7	6
Class of unit	*R*	*R*	*R*	*R*	*R*	*U*	*S*	*R*	*R*	*R*
RS- vector	0	0	0	0	0	—	1	0	0	0

For each of these units a number of one values (n^1) in the unit code bits are indicated. Depending on the proportion of amounts of one and zero values in the code the unit is classified as *R, S* or *U* unit according to the above mentioned rules. In accordance with the results of the classification a *RS*-vector is formed, in which zero values correspond to *R*-units and one values – to *S*-units. *U*-units do not participate in forming the *RS*-vector.

As a result the *RS*-vector takes the value *RS* = 011100001000001000 consisting of 18 bits (the information about two *U*-units was not included in *RS*-vector). Then the *RS*-vector is to be subjected to lossless compression. In the given example the simplest way of compression on the basis of Huffman method is used for illustration (in practice the more effective compression methods are to be used). For this purpose the triads of *RS*-vector bits are taken as elementary characters and their frequency of entering the vector is found. The triad "000" enters the vector twice, the triad "100" – once, the triad "011" – twice. As a result of Huffman method we obtain the following system of uneven prefix codes for the triads of *RS*-vector: "000" ⇒ "11", "100" ⇒ "100", "011" ⇒ "101", "001" ⇒ "0".

As a result of usage of the obtained codes we can have 12 bits vector RS_{com} = 101100011011 instead of the initial triads in 18 bits *RS*-vector. *RS* and RS_{com} vector lengths difference is $\Delta L = 6$. This value expresses the maximum amount of DWM bits, which can be embedded in LUT units in the given example.

In table 3 the values of bits of the initial *RS*-vector and its compressed version RS_{com} are shown. Let us consider the example of addition of the 6-bits DWM – *DWM* = 101010 to the vector RS_{com}. As a result of concatenation of the vector RS_{com} and *DWM* embedded in DWM container we obtain vector *RS** having the similar length as to vector *RS*.

Table 3. Binary vector values

RS	0	1	1	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0
RS_{com}	1	0	1	1	0	0	0	1	1	0	1	1						
DWM													1	0	1	0	1	0
*RS**	1	0	1	1	0	0	0	1	1	0	1	1	**1**	**0**	**1**	**0**	**1**	**0**

Then according to the rules specified in table 1 the LUT-container modification is performed. The aim of this modification is to adapt the LUT-unit classes lying in the embedding path to the value of vector *RS** containing DWM. As we see from table 1 the LUT-unit movement from one class to another is only performed when the values of the corresponding bits of vectors *RS* and *RS** do not coincide. The bits characterizing with the absence of coincidence of such kind are distinguished in table 3. For the LUT-unit codes corresponding to the distinguished bits a Flipping-function is applied and this leads to the movement of these units from one class to another.

In table 4 the values of LUT-unit codes after Flipping-function application to the distinguished units (units, for which the class is to be changed). As a result of these actions the *RS*-vector of units, which are in the embedding path takes a value of the vector *RS** and DWM is embedded into the LUT-container.

Table 4. Values of LUT-unit codes after embedding the DWM in the container

i	1	2	3	4	5	6	7	8	9	10
Unit code	AE0F	0069	56EC	E8FE	2002	0008	8040	EFFF	0FF0	CCCD
n^1	9	4	9	11	2	1	2	15	8	9
Class of unit	*S*	*R*	*S*	*S*	*R*	*R*	*R*	*S*	*U*	*S*
RS-vector	1	0	1	1	0	0	0	1	—	1
i	11	12	13	14	15	16	17	18	19	20
Unit code	0220	7707	0ED7	C0FF	0001	9B94	AABA	A340	4F1E	EE00
n^1	2	9	9	10	1	8	9	5	9	6
Class of unit	*R*	*S*	*S*	*S*	*R*	*U*	*S*	*R*	*S*	*R*
RS-vector	0	1	1	1	0	—	1	0	1	0

4 The Proposed Procedure of DWM Extraction from LUT-Container

Within the frame of the proposed technology the stego-key dedicated for data extraction consists of the following four components:

$$key = (order, classification, RSrule, \Delta L), \quad (3)$$

where *order* – the information defining the order of move around LUT units in the container (embedding path) for DWM embedding or extracting; *classification* – concrete definition of the principle of LUT-unit classification (unit is considered to be *R*-unit if the amount of parts in the unit code is more or less); *RSrule* – the rule of interpretation of *RS*, RS_{com}, *RS** vector bit values: for *R*-units a zero value and for *S*-units one value are fixed in the *RS*-vector and vice versa; ΔL – the difference between the vector *RS* length and its compressed version RS_{com} length.

The succession of actions of the proposed information technology in DWM extracting and recovering the initial value of LUT-container is as follows.

Stage 1. The movement about LUT-container units in the order specified by embedding path occurs. And along with this a binary vector *RS'* is formed on the basis of LUT-unit classification like in the case of embedding the information.

Stage 2. The obtained vector *RS'* structure corresponds to the one of vector *RS** (2) formed in DWM embedding in container. The last ΔL bits of vector *RS'* are DWM with the possible addition to the necessary vector length. At this stage the reading of this bits and DWM obtaining is performed.

Stage 3. The other vector bits are subjected to the procedure of decompression opposite to compression, which is performed at the stage of embedding the information. As a result a binary vector RS_{decom} is formed.

Stage 4. On the basis of information contained in vector RS_{decom} the recovery of original container value is performed. The movement about LUT-units of the container in the order specified by the embedding path and successive review of vector RS_{decom} values are performed for this. In the course of this movement the *U*-

units are ignored and within *R*-units and *S*-units a mutual transition is made by Flipping-function according to table 1.

For the example considered above DWM extraction can be described in the following way. As a result of analysis of LUT-unit codes lying in the embedding path the unit classification is performed and a binary vector *RS'* = 101100011011101010 is formed. From the vector end the $\Delta L = 6$ bits containing DWM "101010" are taking. The rest bits of the vector "101100011011" are subjected to decompression with the usage of the table making the match between the uneven prefix codes of compressed vector and the triads of decompressed vector bits. This leads to creation of binary vector RS_{decom} = 011100001000001000, which coincides with the container vector *RS* before DWM embedding in it. The given vector has the information for the original value recovery of LUT-container. According to the rules of table 1 the LUT-unit classes become the original ones and the container acquires the value it had before DWM embedding.

5 Experimental Research of the Proposed Technology

Aims of the experimental research:

1. to show experimentally that LUT-containers, in which DWMs were embedded with the help of the proposed technology acquire their original value after DWM extraction;
2. to indicate the degree of change of the main characteristics of containers after DWM embedding in them. The main characteristics are considered to be the following ones: a) maximal clock signals frequency expressing the limits of processing speed within the frames of one family of target chips; b) energy consumption of the input-output system of a chip; c) energy consumption of the chip core; d) thermal dissipation of a chip.

Environment for making the experiments. The hardware-software means based on the chips FPGA Altera Cyclone II and CAD Altera Quartus II usage have been developed for the experimental research. The group of scripts performing the interaction with CAD Altera Quartus II for reading and recording the LUT-unit contents was organized in TCL language. The read data processing subsystem itself is realized in language C# within the frame of the platform .Net in accordance with the proposed technology. The estimation of design characteristics mentioned above were carried out by means of CAD Altera Quartus II *Timing Analyzer* (estimation of the limit processing speed) and *Power Play* (estimation of energy consumption and thermal dissipation). The states of container before DWM embedding and after DWM extraction were compared by bit-by-bit analysis of configuration files of the corresponding containers.

The Materials for experiments are 40 FPGA – projects of different complexities and purposes. The hardware complexity of devices within the frames of the given projects varies from 1,2% to 65% resource volume (logic cells, RAM units) of target chip FPGA.

The process of making the experiments. The experiments for each of the researched project consist of the following stages:

1. estimation of the main characteristics (processing speed, energy consumption, thermal dissipation) for the device represented in the form of initial container;
2. embedding the random binary succession of DWM in container;
3. estimation of the main characteristics for a container having the embedded DWM;
4. DWM extraction from the filled container;
5. comparison of configuration files of the final and initial container states.

The results of experimental research. As a result of experiments a complete coincidence of configuration files of the initial containers and the ones obtained after DWM extraction for all of the 40 projects has been established. So due to the features of self-recovery after DWM extraction from containers they take the original form and their characteristics return to the initial values.

In the part of determining the degree of changes of the main container characteristics (after DWM embedding) the following results were obtained. We found that insignificant changes of container characteristics occurs as a result of DWM embedding. Along with this the dependence of this change upon hardware complexity and its structural peculiarities is extremely low. The difference between the changes of characteristics for devices occupying 1,2% and the ones occupying 65% resource volume of the target chip FPGA is hundredths of a percent. On average (for all of the researched 40 projects) the processing speed characteristic change is 0,18%, changes of energy consumption and thermal dissipation characteristics – 0,22%. As we see the change values are within the limits of error value of measurement means.

The technology offered in the paper does not change the mutual connections of LUT units but does change the values of the specific units codes. It is accordingly of interest to learn whether the changes of units codes are able to influence on the FPGA project features under such conditions. Except for the previously considered experiment another experimental research of the influence of the values of LUT units codes on the basic features of FPGA projects has been made with the participation of one of the authors of the given paper. The description of this research and its results are shown in [14]. In the above considered experiment a random binary succession was embedded in various FPGA projects. This led to the fragmentary changes of LUT units codes as in accordance with table 1 the codes of units change only in the two of four cases of possible combinations of values of bits of vectors RS and RS*. In [14] the extreme cases of codes change in all the involved LUT units of FPGA-project were considered. In order to create such kinds of changes the projects of similar structures, but with the LUT units codes containing minimum and maximum amount of one value, respectively, were formed and compared. The projects were chosen to deploy the resources of the corresponding FPGA chip in a maximum way. In table 5 the main results of comparison of such projects are represented. The projects with the minimum and maximum amount of one values are marked in table 5 as P_{min} and P_{max}, respectively. In the given table the following characteristics are represented: maximum clock frequency F; dynamic and static power supply as well as the total

energy consumption of chip core obtained for the indicated clock frequencies; δ – a relative energy consumption change in mass changes of the number of logical values in the codes of LUT units of a project. All the families of chips used in the experiment have the core power supply equal 1,2 V.

Table 5. The results of energy consumption and processing speed estimation

Project	Total core energy consumption (mA)	Dynamic power supply (mA)	Static power supply (mA)
Cyclone II EP2C35F672C6 F = 420,18 MHz			
P_{min}	**699, 17**	629,98	69,19
P_{max}	**701,40**	632,19	69,22
δ	**0,31 %**	0,34 %	0,04 %
Cyclone III EP3C40F780C6 F = 516,26 MHz			
P_{min}	**495,6**	491,49	4,11
P_{max}	**497,37**	493,39	3,98
δ	**0,35 %**	0,38 %	3,16 %
Cyclone III LS EP3CLS70F780C7 F = 516,26 MHz			
P_{min}	**494,6**	482,91	11,69
P_{max}	**496,23**	484,73	11,51
δ	**0,32 %**	0,37%	1,5 %
Cyclone IV E EP4CE30F29C6 F = 600,6 MHz			
P_{min}	**477,42**	470.42	7,00
P_{max}	**479,64**	472,22	7,41
δ	**0,46 %**	0,38 %	5,5 %

The experiment results shown in table 5 demonstrate that in maximum mass change of the number of one values correlation in LUT units:

1. the energy consumption of input output system for chips within a FPGA family is retained at the same level;
2. the chip core energy consumption changes insignificantly within one family mainly due to the dynamic component (maximum value of this change equals 0,46% for chips EP4CE30F29C6 of family Cyclone IV E);
3. the maximum clock frequency expressing the extreme processing speed of devices does not practically change within one FPGA family.

The results show that such important characteristics as productivity and energy consumption are not practically dependent upon the type of Software code used in FPGA-projects with the fixed hardware realization. So we can come to the conclusions that according to the technology proposed in the given paper the DWM embedding in a LUT-container does not substantially impact on the FPGA project characteristics mentioned above.

6 Conclusions

The information technology proposed in the paper allows to perform DWM embedding in containers with LUT-oriented architecture. The technology gives the possibility to recover the initial state of container after DWM extraction from it. The technology is based on the combination of DWM embedding method into LUT-containers proposed in the given paper and the popular Fridrich method oriented at the passive multimedia containers. Unlike Fridrich method the proposed technology:

- uses less complicated Flipping-function taking into account the container peculiarities;
- does not include the discrimination function calculation in order to perform the element classification of container;
- performs information embedding at the level of the container elementary parts (LUT-units) but not at the level of groups of elementary parts.

The degree of the proposed technology effectiveness is of qualitative nature and is expressed in possibility to recover the initial LUT-container after DWM extraction, which was absent before. The proposed technology can be used in developing the hardware and software, which realize digital watermark embedding in computer and control devices created on the LUT-oriented element base (e.g. FPGA or programmable logic integrated circuits with the similar architecture). Such kind of embedding allows to control the device configuration integrity makes the substitution or corruption of configuration impossible, which is of vital importance for critical domains. In addition it allows to control the legitimacy of usage of project information and devices themselves at the different stages of CAD and life cycle: synthesized FPGA project, configuration file FPGA, operating device.

References

1. Shih, F.: Multimedia Security: Watermarking, Steganography, and Forensics. CRC Press, Boca Raton, FL (2013)
2. Cox, I., Miller, M., Bloom, J., Fridrich, J.: Digital Watermarking and Steganography. Morgan Kaufmann Publishers, Amsterdam (2008)
3. Vasu, S., George, S., Deepthi, P.: An Integrity Verification System for Images Using Hashing and Watermarking. Proceedings of the International Conference on Communication Systems and Network Technologies. 85–89 (2012)
4. Coatrieux, G., Hui Huang, Huazhong Shu, Limin Luo, Roux, C.: A Watermarking-Based Medical Image Integrity Control System and an Image Moment Signature for Tampering Characterization. IEEE J. Biomed. Health Inform. 17, 1057–1067 (2013)
5. Sencar, H., Memon, N.: Watermarking and ownership problem. Proceedings of the 5th ACM workshop on Digital rights management – DRM'05. (2005).
6. Arnold, M., Schmucker, M., Wolthusen, S.: Techniques and Applications of Digital Watermarking and Content Protection. Artech House, Boston (2003)
7. Anderson, R.: Security Engineering: A Guide to Building Dependable Distributed Systems, 2nd Edition. Wiley, New York (2008)
8. Huffmire, T.: Handbook of FPGA Design Security. Springer, Dordrecht (2010)

9. Paul, S., Bhunia, S.: Reconfigurable Computing Using Content Addressable Memory for Improved Performance and Resource Usage. In: Proc. Design Automation Conference ACM/IEEE (DAC-2008), 786–791, ACM, Anaheim (2008)
10. Zashcholkin, K., Ivanova, E.: Method of Steganographical Hiding of Information in LUT-Oriented Hardware Containers. Electrical and Computer Systems. 12(88), 83–90 (2013) (in Russian)
11. Zashcholkin, K., Ivanova, E.: Steganography Data Hiding in LUT-Oriented Hardware Containers Method Development. Electrical and Computer Systems. 13(89), 231–239 (2014) (in Russian)
12. Fridrich, J., Goljan, M., Du, R.: Lossless Data Embedding – New Paradigm in Digital Watermarking. EURASIP Journal on Adv. Signal Process. 2002, 185–196 (2002)
13. Goljan, M., Fridrich, J., Du, R.: Distortion-Free Data Embedding for Images. In: Proc. of the 4th International Workshop on Information Hiding (IHW-01), 27–41, USA, Pittsburg (2001)
14. Zashcholkin, K., Kuznetsov, N., Drozd, A.: Research in Key Features of FPGA-projects in Case of Changing the Codes in LUT-blocks. Scientific Herald of Yuriy Fedkovych Chernivtsi National University: Computer Systems and Components. Vol. 5, Issue 2, 82–86 (2014) (in Russian)

Functional Diversity Design of Safety-Related Systems

Ivan Malynyak

Stalenergo LLP
ivanmiros@gmail.com

Abstract. Traditionally, the application of safety voted-groups architectures is a matter of redundancy, where hardware and software components are replicated and become a source of vulnerabilities with decreased system reliability as a whole, therefore necessity of functional diversity design is become essential. Well known diversity approach for similar erroneous results mitigation is widely used, but combined software and hardware techniques to achieve necessary safety system requirements without enlarged implementation of price isn't yet evolved. Avoidance of redundant complexity with limitation the number of channel's internal states could lead to common cause failures reduction and sufficient level of residual risks.

Keywords. fault tolerant architecture, 1oo2D, 2oo3, redundancy, complexity, control systems, diversity, common cause failures

Key terms. Process, Technology, Development, Reliability, SoftwareComponent

1 Introduction

Nowadays technology of control system is believed to be effective and safe, where essential approaches are sufficiently exploited [1]. The current generation of instrumentation and control systems is highly integrated digital complexes which offer better performance, versatility and additional diagnostic capabilities in comparison with aged analog systems. But as far as systems become more safety-critical, all kinds of possible vulnerabilities have to be taken into account.

This paper offers the concept of functional diversity design which is unfolded by three steps from the problem, through different approaches to the implementation. First of all, necessity of latent systematic faults elimination is assumed, where question of misinterpreted functional requirements within the system becomes a main framework for modern applications [2, 3]. Secondly, the approach to cover up to 99.9% faults without overhead in performance and power consumption based on modern surveys is suggested [4]. And finally, new architecture based on reduced system complexity which helps to make up savings in development life cycle stage with higher availability and higher safety in railway industry is proposed.

To be accurate with terminology, the functional diversity design in this paper is not assumed as different physical functions within determined process implementation, but more broadly, where fundamental diverse technology is considered in a way of inherent difference without any commonality in its nature. Even the most comprehensive strategies approach are suffered from the common channel failures and possible external influences or insufficient faults mitigation that can further contribute to the potential concurrent vulnerabilities [2]. Indeed, all mentioned strategies assumed the six different fault management techniques which solves equal project problems in redundant channels, composed of sensors, computers and control elements with purpose to take the process to a de-energized state when predetermined conditions are violated.

As shown in Fig. 1, two types of well-characterized redundant architectures (2oo3 and 1oo2D) have equal functional structure schemes A $\Leftrightarrow$ B $\Leftrightarrow$ C and therefore cycle's synchronization procedures within sub-channels are required. As a result, extra data connection lines between sub-channels are used to establish additional information exchange and projects very often become fed up with wiring L and sophisticated inter-channel redundancy management.

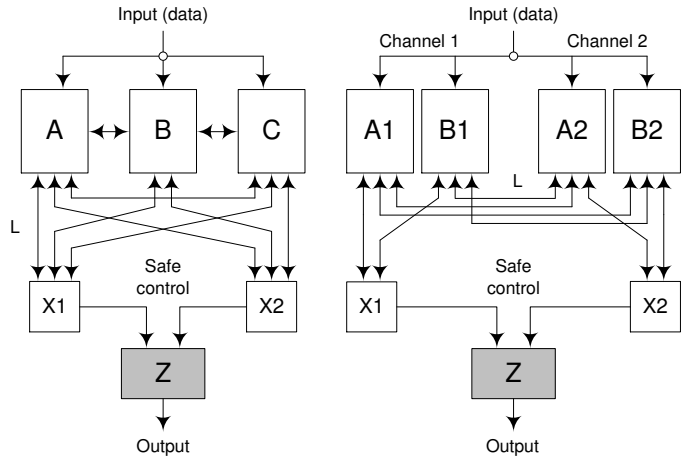


Fig. 1. Two types of well-characterized architectures 2oo3 (left) and 1oo2D (right)

Almost all created inter-channel data than is sent to hardware/software redundant voting mechanism Xi which than implements safe control output function Z. Despite the fact that diversity questions are deeply concerned and surveyed, the biggest issue still lies in assumption of equal channel functional design, where the fundamental problem took their roots [4, 5]. Of course, it's easy to fit out the system with all kinds of diversity techniques, but absence of overall differential approach can't guarantee prolonged operation without common cause vulnerabilities. The possibility of concurrent triggering the latent multi-channel failures is potentially hazardous and it's like a group of man walking simultaneously around a pit with averted eyes.

The way out could be found in “sequential” (or satellite) type of system architecture, where parallel equivalent sub-channels are substituted with diverse ones as shown in Fig. 2. The challenge is to cope with nature evolution, where necessity of redundancy implements as sequential algorithm in embedded bundle of diverse entities [6]. The idea is already has been considered in [2], but existent architectural context is still repeated virtual “OR” logic of distinct sub-channels. Even if functional diversity is achieved through different control laws, elements and distinct functionality, the need of overall agreement among the sub-channels and data synchronization may critically affect output values.

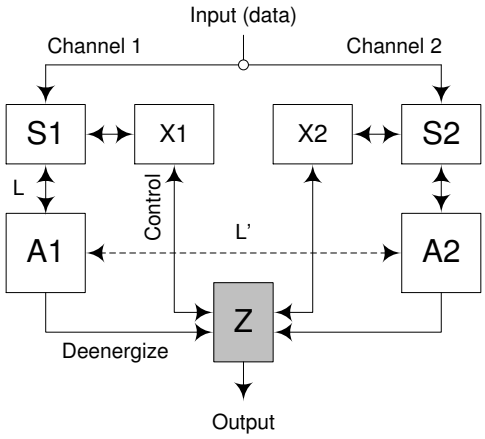


Fig. 2. Sequential architecture of safety redundant control systems 1oo2S

Data lines in sequential architecture 1oo2S are minimized to essential level with safety functions are guaranteed to be executed correctly by de-energize signals. This type of architecture isn’t widespread despite advantages of system simplicity because of general established multichannel parallel approach used to cover safety issues. Nevertheless, such direction may have its future in the light of necessity to drop overall system complexity to acceptable level for commercial use.

2 Redundancy

The challenge of providing redundancy management to meet certain requirements for different application is complicated by the critical constraints of market cost and schedule. Further evolution of redundancy management with its intricate synchronization interfaces and overall complexity were already found to be potentially catastrophic [7]. This approach obviously was costly in a way of wiring and components but from the other hand it’s offered rather simple and attractive macro level “black box” design.

As one could see, the redundancy approach is well known and understandable, where failures of elements could be easily found by comparison. For simple functions it works greatly and reliability decreasing is not matter of consideration. But when dozens of input and output were designed with redundancy management, cost of developing and physical implementation tended to increase exponentially. The result of complexity already known and contained to be in faults propagation which leads to unsafe concurrent failures of two or more identical parallel sub-channels [8].

More than that, the design to achieve necessary safety level leaded out into duplicating functions almost at every step. Firstly, the safe input means at least of two signals to be monitored with safe outputs are based on the same repeating principal. Secondly, to achieve availability requirements the "design shape" just went on adding next "channel 2", but of course it could be more channels. And finally, when all components are successfully placed, the intricately net of data communicating has to be put over.

The first step down from "black boxes" design could be avoiding some needless redundancies. As shown in Fig. 3, 1oo2D architecture is looked safe due to transparent functional scheme. Indeed, all inputs are duplicated and could be verified by other neighbor. Moreover, output Z is drove from two channels by voting mechanisms Xi, with ability to perform safety de-energized state.

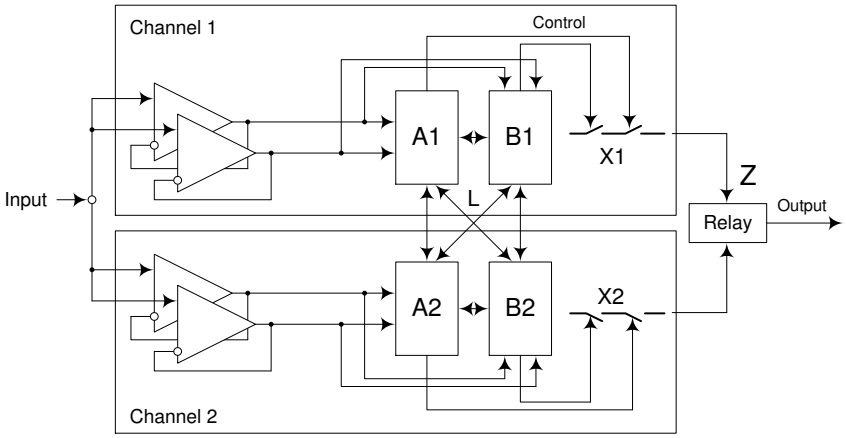


Fig. 3. Decomposition of 1oo2D architecture

On the contrary, the sequential architecture approach could change the whole appearance of system design, where redundancy eliminating is only the first step to be done. As shown in Fig. 4, the point of view lied in the basic sequential approach idea, where elements of system are presented with non-symmetrical faults property. Let's describe it by simplified hardware model, which could be than enlarged on software issues by substitution physical elements with similar software routines in a project.

For example, if "input" is implied as thermocouple, than implementing fixed current with periodical caliber testing would guarantee possible input faults detection. Indeed, the only fault is not detected by this approach is a rare gradual measurement

drift and by using high quality thermocouple with periodical external checkup this could be solved and particular problem due to low possibility of such fault would be eliminated. As a result, a thermocouple with measuring circuits is treated as single element with non-symmetrical types of faults, where most likely defects in circuit break or rapid measurement drift (op-amps or connecting line issues) are easily detected. The non-symmetrical type fault output is treated in the same way, where satellite sub-channel S drives relay Z safely through transformer X with inherently eliminating shortcut failures. The main sub-channel A is connected to S with communication line where control data and both functional integrity is transferred. Each of sub-channels could run safety function by de-energizing transformer in case of discrepancy.

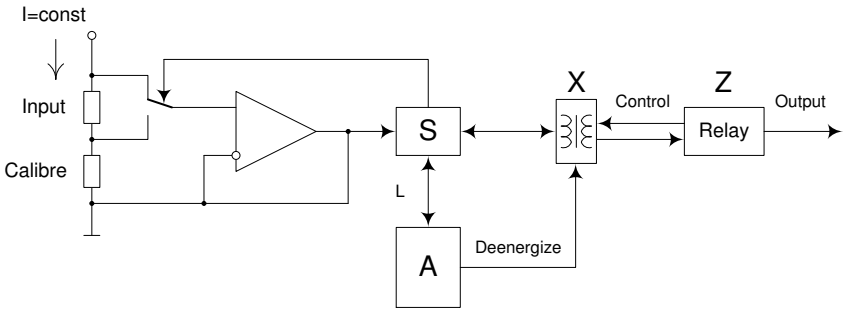


Fig. 4. Simplified model of sequential architecture 1oo1S

As a result, sequential approach is constrained to build the system with strong functional design separation between core algorithm A and routine input/output tasks, whereupon for example the fast and simple satellite sub-channel S is prepared to deal only with input/output X and functional integrity management. The effect would come out of overall system complexity reduction, common cause failures elimination and reliability enlargement due to algorithm simplification with functional diversification.

3 Design of safety-related systems

The main aim of diversity is to perform safety functions achieved by monitoring the behavior of number of sub-channels. The way safety ideas are implemented reminds “fractal” pattern, where general method of reiteration is repeated at every scale of a project. It could be seen in software, where developer has no idea how his source code would be eventually implemented into assembler language. The same thing is happened in hardware, where design is implemented by pure repeating elements based on monolithic integrated circuits without clear understanding of its full functionality. As a result, each level of system decomposition is considered with increasing uncertainty and the principle of aggregation with uncontrollable elements cascading may not be as reliable as it expected.

Finally, the result of uncontrolled “fractal” implementation is come out as necessity of channel diversity due to providing predictable system behavior and is turned to entity multiplication which has to be carefully verified [9]. Hereafter, when system is installed, a well known golden rule is remained - “don’t change anything”.

Evolution of Nature shows that repeated and uncontrolled growth ultimately leads to disaster, just look back on dinosaurs or cancer. So, the main idea of safety may lies in capability of system predicting its behavior with reasonable complexity. For example, for the implication to the brain size it could be said, that decision making mechanism should be “large as you need and as small as you can” [10].

Looking forward at systems design, there is a time point when sufficient data of failures is gathered and overall availability could be increased just due to better maintenance understanding. Indeed, if its clear how light bulbs in mines are breaks and what circumstance is preceded, then particular algorithm with appropriate feedback could lead to limited abundance of lighting control system or even eliminate redundancy at all due to accurate prediction of failures with using certain planned actions.

After thorough revision of prevail railway control systems it’s assumed, that architectures could be transformed into simplified 1oo2S variant, where all inputs and outputs are designed with proper feedback (shown in Fig. 5).

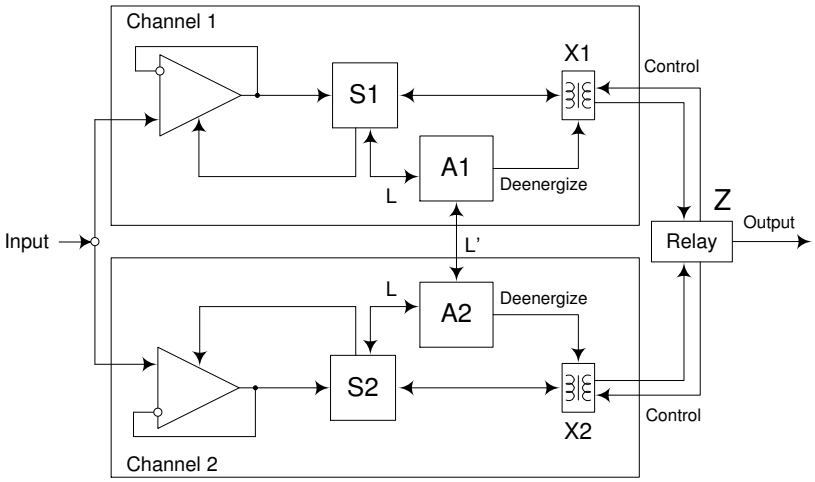


Fig. 5. Decomposition of 1oo2S architecture

For the preliminary evaluation of design safety advantages, it’s assumed that set of proper input signals $p \in P$ or faults $q \in Q$ are used for all sub-channels A,B,C with set of internal conditionals $r \in R$ to implement safely output set Z or hazardous output set Z’.

For 1oo2D architecture $A \Leftrightarrow B$ is assumed where diversity mechanisms within sub-channel could be implemented or not. Working states when safety function operates normally is described by formula:

$$\forall\, r,p:\ Arp=Brp,\ Z\wedge Z'=\varnothing \tag{1}$$

Nonworking de-energized output state where sub-channels' internal conditions are differ because of failure detection:

$$\exists\, r,q,p:\ Arq\neq Brp\vee Arp\neq Brq,\ Z=\varnothing \tag{2}$$

Hazardous state where sub-channels' outputs synchronously being changed by latent failure:

$$\exists\, r,q:\ Arq=Brq,\ Z\wedge Z'\neq\varnothing \tag{3}$$

Sequential architecture 1oo2S is considered to be diverse on design level where each sub-channel is used different technologies, logic and way of responding by nature, therefore equation $A\wedge S\approx\varnothing$ is assumed. Nevertheless, in order to provide necessary safety technique, the idea of artificially divided sub-channels' internal conditionals alphabet is used and proper mapping of normal responses set within each sub-channel is determined [11]:

$$\forall\, r,p:\ Arp\subseteq S_A\wedge Srp\subseteq A_S,\ Z\wedge Z'\approx\varnothing \tag{4}$$

Here the SA is a set of sub-channel A mapping responses which are analyzed in S and vice versa. Hazardous state where sub-channels' outputs synchronously being changed by latent failure are very rare due to functional diversity design:

$$\forall\, r,p,q:\ Arpq\wedge Srpq\approx\varnothing \tag{5}$$

In order to avoid potentially hazardous states, a number of conditions must be met [2]. Let us assume A and S with internal symmetric difference feature and in theory it's possible to reach necessary level of identified errors possibility at 99.9%:

$$\forall\, r,p,q:\ Arpq\Leftrightarrow\neg\, Srpq,\ Z\wedge Z'\approx\varnothing \tag{6}$$

One of the approaches to achieve channels' software symmetric difference could be based on "likely program invariants" idea with substitution invariants property of critical parts for negations ones [12]. For example, if sub-channel A algorithm has CASE operator for "what must be done", than sub-channel's S algorithm to control this part of software should have operator for "what is not supposed to be done in any case".

This paper is regarded only as intention where distinguish functional design specifications of sub-channels is took into account. The formulas (1-6) are correlated with already developed theoretical foundation and mathematical models of multi-version systems [11], where different diversity kinds $r\in R$ are sequentially accumulated in final diverse versions by special mapping of output signal set $Zi\rightarrow Z$.

4 Architecture diversity usage

The problem of common cause failures mitigation basically could be found in four industries (aviation, chemical process, rail transportation and nuclear power plant) that provide high failure-consequence consideration in evaluating the diversity.

The usage of diversity are described in [2] with guidance related to clear examples of diversity (shown in Table 1).

Table 1. Summary of diversity usage

Diversity	Aviation	Railway		Chemical	NPP
Equipment	x		1oo2S	x	x
Design				x	x
Function	x	x		x	x
Human	x	x		x	x
Signal		x		x	x
Software	x	x		x	x

In spite of the intensive researches in area of multi-version systems, the distributed nature of the rail network and the localized action for interlocks and train control, the railway safety management is paid less attention to diversity implementation. By stopping or slowing trains to inhibit access to occupied tracks, railways have a readily accessible safe state by local de-energized “stop” configuration. This failsafe approach based on old all-relay interlocking system resulted in a practical emphasis for identifying faulted conditions and stopping the affected trains until the hazard can be cleared.

In the modern microprocessor interlocks and train control the system diversity usage is vital due to increased speed and shortened time between rolling-stocks. Moreover, unique set of input and output signals for every railway station imposed additional constraints on control systems, which has to be flexible in configuration, safe in operation and cost effective in realization.

The idea of sequential 1oo2S architecture could fill the diversity gap in domains, such as traffic control, where collision avoidance is a key of operational safety but with considerable economic constraints.

5 Conclusions

With respect to size and complexity of modern safety-critical application it could be assumed, that one way to achieve necessary system requirements without enlarged implementation price is functional diversity design optimization. This paper proposes combined software and hardware applying, where overall system complexity would not exceed the uncontrollably high level and would decrease developing time.

Sequential safety architecture 1oo2S combines the benefit of 1oo2D and 2oo3 systems with higher availability and higher safety levels. The price for this innovation lies in additional functional requirements to successfully analyzing integrity of sub-

channels. The additional benefit from such approach could be significant protection against vulnerability of most common cause failures when software design shortcomings and hardware faults are came to light.

References

1. IEC 61508-3:2010: Functional Safety of Electrical/Electronic/ Programmable Electronic Safety-Related Systems – Part 3: Software Requirements
2. Wood, R.; Belles, R., Cetiner, M. & et al. Diversity Strategies for NPP I&C Systems, NUREG/CR-7007 ORNL/TM-2009/302, (2009)
3. Yastrebenetsky, M., Kharchenko, V.: Nuclear Power Plant Instrumentation and Control Systems for Safety and Security. IGI Global, (2014).
4. Yoshikawa, H., Zhang, Z.: Progress of Nuclear Safety for Symbiosis and Sustainability: Advanced Digital Instrumentation, Control and Information Systems for Nuclear Power Plants. Springer, Japan, (2014)
5. Avizienis, A., Laprie, J.-C., Randell, B.: Fundamental Concepts of Dependability. Research Report No 1145, LAAS-CNRS, (2001)
6. Mukai, Y., Tohma, Y.: A Method for the Realization of Fail-Safe Asynchronous Sequential Circuits. IEEE Trans. Computer. 23(7), 736–739, (1974)
7. Boykin, J., Thibodeau, J., Schneider, H.: Evolution of Shuttle Avionics Redundancy Management/Fault Tolerance. Space Shuttle Technical Conference, NASA Conference Publication 2342. Part 1, Johnsons Space Center, Texas, pp.1–18, (1983)
8. Madden, W., Rone, K.: Design, Development, Integration: Space Shuttle Primary Flight Software System. Communications of the ACM 27(9), 914–925, (1984)
9. Astrom, K., Murray, R.: Feedback Systems: an Introduction for Scientists and Engineers. Princeton Univ. Press, (2008)
10. Davidson, I.: As Large as You Need and as Small as You Can: Implications of the Brain Size of Homo Floresiensis. In Schalley, A., Khlentzos D.: Mental States. V.1: Evolution, function, nature, pp. 35–42, (2007)
11. Kharchenko, V., Siora, A., Sklyar, V.: Multi-Diversity Versus Common Cause Failures: FPGA-Based Multi Version NPP I&C systems, Proceeding of the 76th conference NPIC&HMIT, Las-Vegas, Nevada, USA, (2010)
12. Sahoo, S., Li, M., Ramachandran, P., Adve, S., Adve,V., Zhou, Y.: Using Likely Program Invariants to Detect Hardware Errors. In Conf. Dependable Systems and Networks – DSN, pp. 70–79, (2008)

Computer's Analysis Method and Reliability Assessment of Fault-Tolerance Operation of Information Systems

Igor P. Atamanyuk[1], Yuriy P. Kondratenko[2]

[1] Mykolaiv National Agrarian University, Commune of Paris str. 9,
54010 Mykolaiv, Ukraine
atamanyukip@mnau.edu.ua
[2] Petro Mohyla Black Sea State University, 68th Desantnykiv Str. 10,
54003 Mykolaiv, Ukraine
yuriy.kondratenko@chdu.edu.ua

Abstract. In this paper there was obtained an calculation method of the assessment of the probability of fail-safe operation of information systems in the future instants of time. The method is based on the algorithm for modeling a posteriori nonlinear random sequence of change of values of the controlled parameters which is imposed a limitation of belonging to a certain range of possible values. The probability of fail-safe operation is defined as the ratio of the number of realizations that fell in the allowable range to the total number of them, formed as a result of the numerical experiment. The realization of an a posteriori random sequence is an additive mixture of optimal from the point of view of mean-square nonlinear estimate of the future value of the parameter analyzed and of the value of a random variable, which may not be predicted due to the stochastic nature of the parameters. The model of a posteriori random sequence is based on the Pugachev's canonical expansion. The calculation method offered does not impose any significant constraints on the class of random sequences analyzed (linearity, stationarity, Markov behavior, monotoneness, etc.).

Keywords. calculation method, random sequence, canonical decomposition, estimation, probability.

Key Terms. computation, mathematical model.

1 Introduction

One of the most important problems that arises constantly in the process of information and communication systems service is the problem of the estimation of system reliability with the following making decision about the possibility of its further exploitation on the basis of the information about possible failures in the future [1-5]. The problem becomes especially important in the case when information system is used for the management of the objects that relate to the class of critical or dangerous and under the threat of accident objects (aircraft, sea mobile objects, nuclear power stations, chemical industry plants etc.) [6-8]. The forecasting of failures is the primary

stage of the providing of the dependability of the systems of such class. However, nowadays the problem of failures forecasting (in overwhelming majority cases) is solved by means of informal methods and the decision about the possibility of dependable exploitation of the corresponding information system is made on the basis [9-11]:

– qualitative and quantitative estimation of its current state;
– the experience of the exploitation of the given and analogous information systems.

As far as information and communication systems (ICT) become more complicated and also as far as there is the growth of requirements to the probability of their reliable work, informal methods of making decision are becoming less and less effective. Hence, the usage of stricter approaches to the estimation of the dependability and reliability of ICTs functioning based on the quantitative estimations of future state of experimental information systems is very important.

2 Statement of the problem

Accuracy of the obtained solutions for correspondent class of problems and the time during which given solutions can be obtained are the most universal parameters characterizing the quality of information system functioning. Stated parameters (accuracy and time) have stochastic character because of the presence of inner defects of the system and uncontrolled external destabilizing factors. That's why the problem of the forecasting control of information system reliability can be formulated in the following way.

Without the restriction of the generality of the next calculations let us assume that the state of the information system in exhaustive way is determined with scalar parameter X (accuracy or time of problem solution). The change of the values of the parameter X in discrete range of points $t_i, i=\overline{1,I}$ is described by random sequence $\{X\}=X(i),\ i=\overline{1,I}$. The values of the parameter X must satisfy the condition

$$a<x(i)<b,\ i=\overline{1,I} \tag{1}$$

In the case of the crossing by parameter X the limits of acceptable area [a; b] the failure of information system in the process of its functioning is registered. The state of the system is controlled periodically in discrete moments of time $t_\mu,\ \mu=\overline{1,k}$ by measurement of the values $x(\mu),\ \mu=\overline{1,k}$ of the parameter X where $x(\mu),\ \mu=\overline{1,k}$ is the realization of the random sequence $\{X\}$ in the cut set $t_\mu,\ \mu=\overline{1,k}$. It is evident that for the segment of the time $[t_1;t_k]$ the inequation $a<x(\mu)<b, \mu=\overline{1,k}$ must be correct. Otherwise as it follows from (1) under

$$x(i)<a,\ i=\overline{1,I}$$

or

$$x(i) > b,\ i = \overline{1,I}$$

on the interval of examination $[t_1; t_k]$ the failure would take place that would lead to the suspension of the process of information system functioning.

The statement of the problem can be formulated in the following way: on the basis of stated (measured) information about current values of the parameter X on the time interval $[t_1; t_k]$ the conclusion about the usability of the information system to exploitation in future moments of time $t_i, i = \overline{k+1, I}$ must be made.

Similarly the problem of providing of dependable functioning and forecasting of the state of the technical systems or objects to the structure of which information systems or management-information systems enter in the form of components can be formulated. Herewith the dependability of functioning and usability of such complicated systems and objects certainly depend on reliability and fail-safety of all their components..

3 Solution

Given that the value of the controlled parameter X changes randomly within the forecast region, the probability of fail-safe operation becomes an exhaustive feature of the safety of functioning of the information system examined.

$$P^{(k)}(I) = P\{a < X^{(k)}(i) < b, i = \overline{k+1, I} / x(\mu), \mu = \overline{1,k}\}. \quad (2)$$

The problem is thus reduced to the determination of the probability of non-falling of the realization of the a posteriori random sequence $X^{(k)}(i / x(\mu), \mu = \overline{1,k}), i = \overline{k+1, I}$ outside the limits of the permissible region $[a; b]$.

In [12], [13] there was proposed an approach to the estimation of likelihood (2) through multiple statistical modeling of possible extensions $x_l(i), i = \overline{k+1, I}, l = \overline{1, L}$ of the random sequence analyzed $\{X\}$ in the forecast region, verification for each realization of condition (1) and calculation as a result of the required estimation experiment $P^{*(k)}(I) = n/L$ (n is the number of successes). In this method, its canonical expansion [14] within the range of points analyzed is used as a model of the random sequence $t_i, i = \overline{1, I}$:

$$X(i) = M\left[X(i)\right] + \sum_{\nu=1}^{i} V_\nu \phi_\nu(i), i = \overline{1, I}, \quad (3)$$

where $V_\nu, \nu = \overline{1, I}$ - random coefficients: $M[V_\nu] = 0, M[V_\nu V_\mu] = 0$ for $\nu \neq \mu, M[V_\nu^2] = D_\nu$;

$\phi_\nu(i), i, \nu = \overline{1, I}$ - nonrandom coordinate function: $\phi_\nu(\nu) = 1, \phi_\nu(i) = 0$ under $\nu > i$.

Elements of the canonical representation (3) are defined by the following recursions:

$$V(i) = X(i) - M[X(i)] - \sum_{\nu=1}^{i-1} V_\nu \varphi_\nu(i), i = \overline{1, I}, \tag{4}$$

$$D_i = M[X^2(i)] - \{M[X(i)]\}^2 - \sum_{\nu=1}^{i-1} D_\nu \varphi_\nu^2(i), i = \overline{1, I}; \tag{5}$$

$$\begin{aligned} \varphi_\nu(i) = \frac{1}{D_\nu} \{ & M[X(\nu)X(i)] - M[X(\nu)]M[X(i)] - \\ & - \sum_{j=1}^{\nu-1} D_j \varphi_j(\nu)\varphi_j(i)\}, \nu = \overline{1, I}, i = \overline{\nu, I}. \end{aligned} \tag{6}$$

Tipping in the expression (3) of known values $X(\mu) = x(\mu), \mu = \overline{1, k}$ converts the a priori random sequence into the a posteriori one:

$$X^{(k)}(i) = m_x^{(k)}(i) + \sum_{\nu=k+1}^{i} V_\nu \phi_\nu(i), i = \overline{k+1, I}, \tag{7}$$

where $m_x^{(k)}(i)$ - linear optimal, by the criterion of mean square minimum of prediction error, estimate of the future value of the random sequence $\{X\}$ at the point t_i according to the known initial values of k.

Expressions for finding $m_x^{(k)}(i)$ have two equivalent forms of notation

$$m_x^{(\mu)}(i) = \begin{cases} M[X(i)], \text{ if } \mu=0, \ i=\overline{1,I}; \\ m_x^{(\mu-1)}(i) + [x(\mu) - m_x^{(\mu-1)}(\mu)]\phi_\mu(i), \ \mu=\overline{1,k}, \ i=\overline{\mu+1,I}; \end{cases} \tag{8}$$

or

$$m_x^{(k)}(i) = M[X(i)] + \sum_{j=1}^{k} (x(\mu) - M[X(\mu)]) f_\mu^{(k)}(i), \ i = \overline{k+1, I}; \tag{9}$$

$$f_\mu^{(k)}(i) = \begin{cases} f_\mu^{(k-1)}(i) - f_\mu^{(k-1)}(k)\phi_k(i), \ \mu \le k\text{-}1; \\ \phi_k(i), \ \mu=k; \end{cases} \tag{10}$$

Formation of possible extensions of random sequence $\{X\}$ by the expression (7) is to compute estimates $m_x^{(k)}(i), i = \overline{k+1, I}$, generating one of the known methods of statis-

tical modeling of values of independent random coefficients $V_\nu, \nu = \overline{k+1, I}$ with the required distribution law and transforming of the values obtained by the coordinate functions $\varphi_\nu(i), i, \nu = \overline{k+1, I}$.

The calculation method of forecasting of fail-safe operation of information systems on the basis of the model (7) covers a fairly wide class of random sequences (non-markovian, non-stationary, non-monotonic, etc.), but this representation of an a posteriori random sequence is optimal only within the framework of linear stochastic properties, thus reducing significantly the accuracy of prediction of random sequences, which have non-linear links.

The clearing of this trouble is possible through the use on the basis of estimation method of the probability of fail-safe operation of an information system of nonlinear canonical expansion of the random sequence [15], changing values of the parameter controlled:

$$X(i) = M[X(i)] + \sum_{\nu=1}^{i} \sum_{\lambda=1}^{N-1} V_\nu^{(\lambda)} \phi_{1\nu}^{(\lambda)}(i), i = \overline{1, I}. \tag{11}$$

Elements of the expansion (11) are determined by the following recursions:

$$V_\nu^{(\lambda)} = X^\lambda(\nu) - M[X^\lambda(i)] - \sum_{\mu=1}^{\nu-1} \sum_{j=1}^{N-1} V_\mu^{(j)} \varphi_{\lambda\mu}^{(j)}(\nu) - \sum_{j=1}^{\lambda-1} V_\nu^{(j)} \varphi_{\lambda\nu}^{(j)}(\nu), \ \nu = \overline{1, I}; \tag{12}$$

$$\begin{aligned} D_\lambda(\nu) = M[\{X(\nu) - M[X(\nu)]\}^{2\lambda}] - \sum_{\mu=1}^{\nu-1} \sum_{j=1}^{N-1} D_j(\mu) \{\phi_{\lambda\mu}^{(j)}(\nu)\}^2 - \\ - \sum_{j=1}^{\lambda-1} D_j(\nu) \{\phi_{\lambda\nu}^{(j)}(\nu)\}^2, \nu = \overline{1, I}; \end{aligned} \tag{13}$$

$$\begin{aligned} \phi_{h\nu}^{(\lambda)}(i) = \frac{1}{D_\lambda(\nu)} \{M[X^\lambda(\nu) X^h(i)] - M[X^\lambda(\nu)] M[X^h(i)] - \\ - \sum_{\mu=1}^{\nu-1} \sum_{j=1}^{N-1} D_j(\mu) \phi_{\lambda\mu}^{(j)}(\nu) \phi_{h\mu}^{(j)}(i) - \sum_{j=1}^{\lambda-1} D_j(\nu) \phi_{\lambda\nu}^{(j)}(\nu) \phi_{h\nu}^{(j)}(i)\}, \lambda = \overline{1, h}, \nu = \overline{1, i}. \end{aligned} \tag{14}$$

In the canonical expansion (11) the random sequence $\{X\}$ is represented in the range of points analyzed $t_i, i = \overline{1, I}$ via $N-1$ the arrays $\{V^{(\lambda)}\}, \lambda = \overline{1, N-1}$ of uncorrelated centered random coefficients. $V_i^{(\lambda)}, \lambda = \overline{1, N-1}, i = \overline{1, I}.$ These coefficients contain information on the values $X^\lambda(i), \lambda = \overline{1, N-1}, i = \overline{1, I}$, and the coordinate functions $\phi_{h\nu}^{(\lambda)}(i), \lambda, h = \overline{1, N-1}, \nu, i = \overline{1, I}$ describe probabilistic links of the order $\lambda + h$ between the sections t_ν and $t_i, \nu, i = \overline{1, I}.$

Block-diagram of the procedure for calculating the parameters of the canonical decomposition is shown in Fig. 1.

The concretization of values $X^{\lambda}(\mu)=x^{\lambda}(\mu),\lambda=\overline{1,N-1},\mu=\overline{1,k}$ allows to move from the a priori random sequence (11) to the a posteriori one:

$$X^{(k)}(i)=m_x^{(k,N-1)}(1,i)+\sum_{\nu=k+1}^{i}\sum_{\lambda=1}^{N-1}V_{\nu}^{(\lambda)}\phi_{1\nu}^{(\lambda)}(i),i=\overline{1,I}. \tag{15}$$

The expression $m_x^{(k,l)}(1,i)=M[X(i)/x^{\nu}(j),j=\overline{1,k},\nu=\overline{1,N-1}]$ is the conditional expectation of a random sequence providing that values $x^{\nu}(j),\nu=1,\overline{N-1},j=\overline{1,k}$ are known and the process analyzed is fully specified by the discretized moment functions $M[X^{\lambda}(\nu)],M[X^{\lambda}(\nu)X^{h}(i)],=\lambda,h=\overline{1,N-1},\nu,i=\overline{1,I}.$

The computing algorithm $m_x^{(k,l)}(1,i)=M[X^1(i)/x^{\nu}(j),j=\overline{1,k},\nu=\overline{1,N-1}]$ has two equivalent forms of notation [16]:

$$m_x^{(\mu,l)}(h,i)=\begin{cases}M[X^h(i)],\quad \mu=0;\\ m_x^{(\mu,l-1)}(h,i)+(x^l(\mu)-m_x^{(\mu,l-1)}(l,\mu))\phi_{h\mu}^{(l)}(i),l\neq 1,\\ m_x^{(\mu-1,N-1)}(h,i)+(x(\mu)-m_x^{(\mu-1,N-1)}(1,\mu))\phi_{h\mu}^{(1)}(i),l=1\end{cases} \tag{16}$$

or

$$m_x^{(k,N-1)}(1,i)=M[X(i)]+\sum_{j=1}^{k}\sum_{\nu=1}^{N-1}x^{\nu}(j)F_{((j-1)(N-1)+\nu)}^{(k(N-1))}((i-1)(N-1)+1), \tag{17}$$

where

$$F_{\lambda}^{(\alpha)}(\xi)=\begin{cases}F_{\lambda}^{(\alpha-1)}(\xi)-F_{\lambda}^{(\alpha-1)}(\alpha)\gamma_k(i),\ \ \lambda\le\alpha\text{-}1;\\ \gamma_{\alpha}(\xi),\ \lambda{=}\alpha;\end{cases} \tag{18}$$

$$\gamma_{\alpha}(\xi)=\begin{cases}\varphi_{1,[\alpha/(N-1)]+1}^{(\mathrm{mod}_{N-1}(\alpha))}([\alpha/(N-1)]+1),\ \text{for}\ \xi\le k(N-1);\\ \varphi_{1,[\alpha/(N-1)]+1}^{(\mathrm{mod}_{N-1}(\alpha))}(i),\ \text{if}\ \xi{=}(i\text{-}1)(N\text{-}1)+1.\end{cases} \tag{19}$$

The simulation procedure of the a posteriori random sequence (15) assumes that densities of random coefficients $V_i^{(\lambda)},\lambda=\overline{1,N-1},i=\overline{1,I}$ are known. The simplest and the most effective solution to the problem of determining these one-dimensional densities is to use nonparametric parse type estimates [17]. Together with this the estimate of the required density of distribution $f(V_i^{(\lambda)})$ of the random variable $V_i^{(\lambda)}$ according to L of its realization $\mathrm{v}_{i,l}^{(\lambda)},l=\overline{1,L}$ is represented as

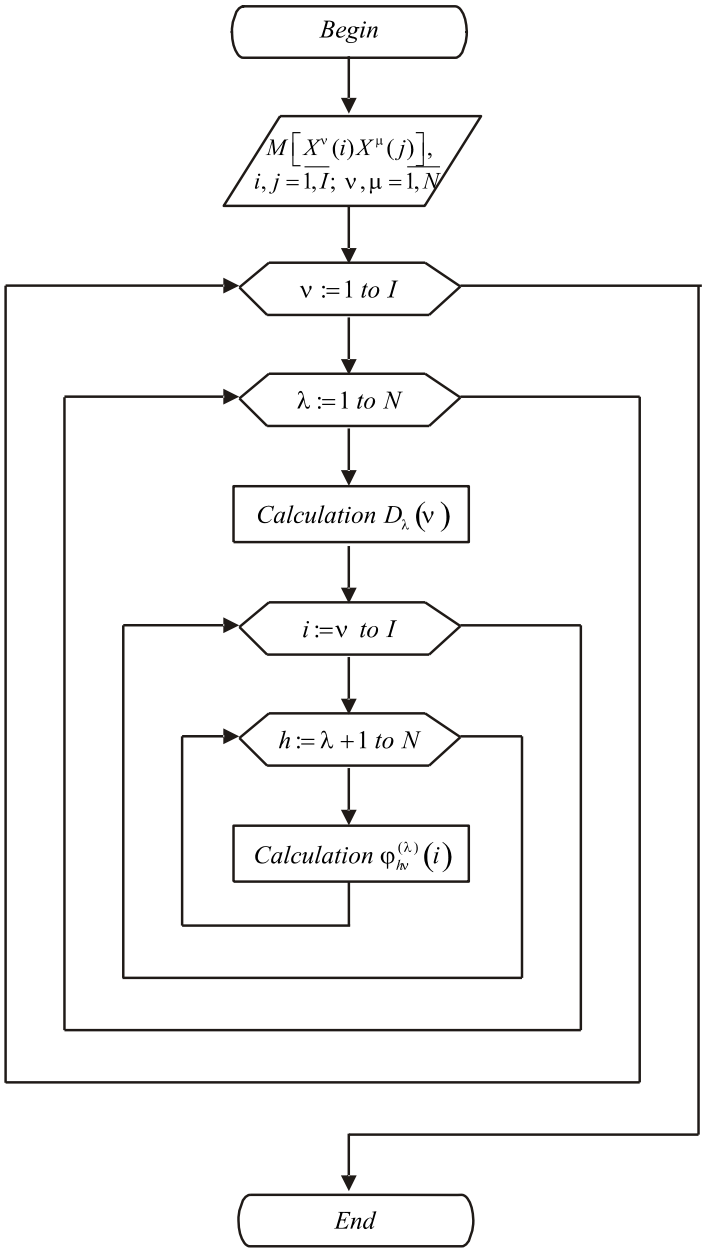


Fig. 1. Block-diagram of the procedure for calculating the parameters of the canonical decomposition (11).

$$f_L(V_i^{(\lambda)}) = \frac{1}{dL}\sum_{l=1}^{L} g(u_l), \tag{20}$$

where $u_l = d^{-1}(v_i^{(N)} - v_{i,l}^{(N)})$, $g(u_l)$ - certain weight function (kernel);

d - constant (blur coefficient).

The estimate (20) at all points of the determination region is obtained unbiased, consistent and uniformly converges on the desired distribution density $f(V_i^{(\lambda)})$ with probability one, if the weight function fulfils the condition

$$g(u) \ge 0; \quad \sup_u |g(u)| < \infty; \quad \lim_{u\to\pm\infty} |ug(u)| = 0; \quad \int_{-\infty}^{+\infty} g(u)du = 1. \tag{21}$$

The constant d is selected depending on the number of observations subject to the conditions

$$d > 0; \quad \lim_{L\to+\infty} d(L) = 0; \quad \lim_{L\to+\infty} d(L)L = \infty. \tag{22}$$

When selected as the kernel function $g(u)$ the uniform density distribution blur coefficient is determined on the basis of the correlation

$$d = 0,5\sup_l \left| v_{i,l}^{(N)} - v_{i,l-1}^{(N)} \right|, v_{i,l}^{(N)} > v_{i,l-1}^{(N)}, l = \overline{2,L}. \tag{23}$$

Thus, the offered calculation method of polynomial predictive control of fail-safe operation of information systems consists of the following stages:

- construction on the basis of the known a priori information $M[X^{\lambda}(\nu)], M[X^{\lambda}(\nu)X^{h}(i)], = \lambda, h = \overline{1, N-1}, \nu, i = \overline{1, I}.$ of the canonical expansion (11) of the random sequence of change of the controlled parameter X;
- determination from the formula (16) or (17) the values $m_x^{(k,l)}(1,i) = M[X(i)/x^{\nu}(j), j = \overline{1,k}, \nu = \overline{1,N-1}]$ of the conditional expectation of the random sequence analyzed within the forecast region $[t_{k+1}..t_I]$ using the known values $x^{\nu}(j), \nu = 1, \overline{N-1}, j = \overline{1,k}$ on the observation interval $[t_1..t_k]$;
- multiple simulation of values of random coefficients $V_i^{(\lambda)}, i = \overline{k+1, I}, \lambda = \overline{1, N-1}$ under the distribution law (20) and formation using the expression (15) of the set of possible extensions of the realization of the random sequence within the forecast range $[t_{k+1}..t_I]$;
- verification of conditions of non-crossing by the paths obtained of the boundaries of the admissible region $[a;b]$ of the controlled parameter change X and determination of the estimate of the probability of fail-safe operation of the in-

formation system as the ratio of the number of successes to the total number of the experiments conducted.

Increasing the reliability of the estimate of the probability of fail-safe operation on the basis of the model (15) compared to (7) is achieved by using nonlinear stochastic properties of the random sequence analyzed: there rises the accuracy of determination of conditional expectation and reliability of possible paths of a random sequence in the forecast region through the use in the process of simulation of an additional array of random coefficients $V_i^{(\lambda)}, i=\overline{\mathrm{k}+1,I}, \lambda=\overline{2,N-1}$. The gain in accuracy can be estimated using the expression:

$$e_{[a,b]}^{(k)}=\frac{\left|m_x^{(k,N-1)}(1,i)-m_x^{(k)}(i)\right|+\{\sum_{j=1}^{k}\sum_{\nu=1}^{N-1}D_\nu(j)(\beta_{1j}^{(\nu)}(i))^2-\sum_{j=1}^{k}D_j\phi_\nu^2(i)\}^{1/2}}{b-a}. \quad (24)$$

Let a random sequence $\{X\}$ in the discrete row of points t_i, $i=\overline{1,I}$ is set by instant functions: $M\left[X^{\xi_l}\left(i-p_{l-1}\right)X^{\xi_{l-1}}\left(i-p_{l-2}\right)...X^{\xi_2}\left(i-p_1\right)X^{\xi_1}\left(i\right)\right]$, $\sum_{j=1}^{l}\xi_j\le N$, $p_j=\overline{1,i-1}$, $i=\overline{1,I}$. Decomposition of random sequence $\{X\}$ looks like [18],[19]:

$$X\left(i\right)=M\left[X\left(i\right)\right]+\sum_{\nu=1}^{i-1}\sum_{\xi_1^{(1)}=1}^{N-1}V_{\xi_1^{(1)}}\left(\nu\right)\varphi_{\xi_1^{(1)}}^{(1)}\left(\nu,i\right)+V_1\left(i\right)+$$

$$+\sum_{\nu=1}^{i-1}\sum_{l=2}^{M(\nu)}\sum_{p_1^{(l)}=1}^{p_1^{'(l)}}\cdots\sum_{p_{l-1}^{(l)}=p_{l-2}^{(l)}+1}^{p_{l-1}^{'(l)}}\sum_{\xi_1^{(l)}=1}^{\xi_1^{'(l)}}\cdots\sum_{\xi_l^{(l)}=1}^{\xi_l^{'(l)}}V_{p_1^{(l)}\ldots p_{l-1;}^{(l)}\xi_1^{(l)}\ldots\xi_l^{(l)}}\left(\nu\right)\times \quad (25)$$

$$\times\varphi_{p_1^{(l)}\ldots p_{l-1;}^{(l)}\xi_1^{(l)}\ldots\xi_l^{(l)}}^{(1)}\left(\nu,i\right),\ i=\overline{1,I},$$

where

$$M\left(\nu\right)=\begin{cases}\nu, & \text{if} \quad \nu<N,\\ N, & \text{if} \quad \nu\ge N;\end{cases}$$

$$p_j^{'(l)}=\begin{cases}0, & \text{if} \quad j\ne\overline{1,l-1} \text{ or } l=1,\\ \nu-l+j, & \text{if} \quad j\ne\overline{1,l-1}, l>1;\end{cases}$$

$$\xi_\mu^{'(l)}=N-l+\mu-\sum_{j=1}^{\mu-1}\xi_j^{(l)},\ \mu=\overline{1,l}.$$

The casual coefficients of canonical presentation (9) are defined by the expressions:

$$V_{\alpha_1}(\nu)=X^{\alpha_1}(\nu)-M\left[X^{\alpha_1}(\nu)\right]-\sum_{\lambda=1}^{\nu-1}\sum_{\xi_1^{(1)}=1}^{N}V_{\xi_1^{(1)}}(\lambda)\varphi^{(1)}_{\xi_1^{(1)}}(\lambda,\nu)-\sum_{\xi_1^{(1)}=1}^{\alpha_1-1}V_{\xi_1^{(1)}}(\nu)$$
$$\times\varphi^{(\alpha_1)}_{\xi_1^{(1)}}(\nu,\nu)-\sum_{\lambda=2}^{\nu-1}\sum_{l=2}^{M(\lambda)}\sum_{p_1^{(l)}=1}^{p_1^{'(l)}}\ldots\sum_{p_{l-1}^{(l)}=p_{l-2}^{(l)}+1}^{p_{l-1}^{'(l)}}\sum_{\xi_1^{(l)}=1}^{\xi_1^{'(l)}}\ldots\sum_{\xi_l^{(l)}=1}^{\xi_l^{'(l)}}V_{p_1^{(l)}\ldots p_{l-1;}^{(l)}\xi_1^{(l)}\ldots\xi_l^{(l)}}(\lambda)\times \quad (26)$$
$$\times\varphi^{(\alpha_1)}_{p_1^{(l)}\ldots p_{l-1;}^{(l)}\xi_1^{(l)}\ldots\xi_l^{(l)}}(\lambda,\nu),\ \nu=\overline{1,I}.$$

The coefficients $V_{\beta_1,\ldots\beta_{n-1};\alpha_1,\ldots\alpha_n}(\nu)$ which contain information about the values $X^{\alpha_n}(\nu-\beta_{n-1})\ldots X^{\alpha_1}(\nu)$ are calculated as

$$V_{\beta_1,\ldots\beta_{n-1};\alpha_1,\ldots\alpha_n}(\nu)=X^{\alpha_n}(\nu-\beta_{n-1})\ldots X^{\alpha_1}(\nu)-M\left[X^{\alpha_n}(\nu-\beta_{n-1})\ldots X^{\alpha_1}(\nu)\right]-$$
$$-\sum_{\lambda=1}^{\nu}\sum_{\xi_1^{(1)}=1}^{N-1}V_{\xi_1^{(1)}}(\lambda)\varphi^{(\beta_1,\ldots\beta_{n-1};\alpha_1,\ldots\alpha_n)}_{\xi_1^{(1)}}(\lambda,\nu)-$$
$$-\sum_{\lambda=1}^{\nu-1}\sum_{l=2}^{M(\lambda)}\sum_{p_1^{(l)}=1}^{p_1^{'(l)}}\ldots\sum_{p_{l-1}^{(l)}=p_{l-2}^{(l)}+1}^{p_{l-1}^{'(l)}}\sum_{\xi_1^{(l)}=1}^{\xi_1^{'(l)}}\ldots\sum_{\xi_l^{(l)}=1}^{\xi_l^{'(l)}}V_{p_1^{(l)}\ldots p_{l-1;}^{(l)}\xi_1^{(l)}\ldots\xi_l^{(l)}}(\lambda)\times$$
$$\times\varphi^{(\beta_1,\ldots\beta_{n-1};\alpha_1,\ldots\alpha_n)}_{p_1^{(l)}\ldots p_{l-1;}^{(l)}\xi_1^{(l)}\ldots\xi_l^{(l)}}(\lambda,\nu)- \quad (27)$$
$$-\sum_{l=2}^{n-1}\sum_{p_1^{(l)}=1}^{p_1^{'(l)}}\ldots\sum_{p_{l-1}^{(l)}=p_{l-2}^{(l)}+1}^{p_{l-1}^{'(l)}}\sum_{\xi_1^{(l)}=1}^{\xi_1^{'(l)}}\ldots\sum_{\xi_l^{(l)}=1}^{\xi_l^{'(l)}}V_{p_1^{(l)}\ldots p_{l-1;}^{(l)}\xi_1^{(l)}\ldots\xi_l^{(l)}}(\nu)\times$$
$$\times\varphi^{(\beta_1,\ldots\beta_{n-1};\alpha_1,\ldots\alpha_n)}_{p_1^{(l)}\ldots p_{l-1;}^{(l)}\xi_1^{(l)}\ldots\xi_l^{(l)}}(\nu,\nu)-$$
$$\sum_{p_1^{(n)}=1}^{p_1^{*(n)}}\ldots\sum_{p_{l-1}^{(n)}=p_{l-2}^{(n)}+1}^{p_{n-1}^{*(n)}}\sum_{\xi_1^{(n)}=1}^{\xi_1^{*(n)}}\ldots\sum_{\xi_n^{(n)}=1}^{\xi_n^{*(n)}}V_{p_1^{(n)}\ldots p_{n-1;}^{(n)}\xi_1^{(n)}\ldots\xi_n^{(n)}}(\nu)\varphi^{(\beta_1,\ldots\beta_{n-1};\alpha_1,\ldots\alpha_n)}_{p_1^{(n)}\ldots p_{n-1;}^{(n)}\xi_1^{(n)}\ldots\xi_n^{(n)}}(\nu,\nu).$$

In (27) parameters $p_1^{*(n)},\ldots,p_{n-1}^{*(n)};\xi_1^{*(n)},\ldots,\xi_n^{*(n)}$ are calculated by the following expressions:

$$p_j^{*(n)} = \begin{cases} \beta^*{}_\mu, & \text{if} \quad \mu = 1,\ p_{\mu-1}^{(n)} = \beta^*{}_\mu,\ \mu = \overline{2,n}, \\ \nu - l + \mu, & \text{if}\ p_{\mu-1}^{(n)} = \beta^*{}_{\mu-1},\ \mu = \overline{2,n}. \end{cases} \tag{28}$$

$$\xi_j^{*(n)} = \begin{cases} \alpha^*{}_i, & \text{if} \quad i = 1,\ \xi_j^{(n)} = \alpha^*{}_{i-1}, i = \overline{2,n}; \\ N - n + i - \sum\limits_{j=1}^{i=1} \xi_j^{(n)}, & \text{if} \quad \xi_j^{(n)} \neq \alpha^*{}_{i-1}, i = \overline{2,n}. \end{cases} \tag{29}$$

The values $\beta^*{}_\mu,\ \mu = \overline{1,n-1}; \alpha^*{}_i,\ i = \overline{1,n},$ are the indexes of casual coefficient $V_{\beta^*{}_1 \ldots \beta^*{}_{n-1}; \alpha^*{}_1 \ldots \alpha^*{}_n}(\nu)$ which proceeds $V_{\beta_1 \ldots \beta_{n-1}; \alpha_1 \ldots \alpha_n}(\nu)$ in canonical decomposition (25) for the moment of time t_ν :

1. $\beta^*{}_\mu = \beta_\mu, \mu = \overline{1,n-1};\ \alpha^*{}_i = \alpha_i, i = \overline{1,k-1};\ \alpha^*{}_k = \alpha_k - 1;$

$\alpha^*{}_j = N - n + j - \sum\limits_{m=1}^{j-1} \alpha^*{}_m,\ j = \overline{k+1,n};$ if $\alpha_k > 1, \alpha_j = 1,\ j = \overline{k+1,n};$

2. $\beta^*{}_\mu = \beta_\mu, \mu = \overline{1,k-1};\ \beta^*{}_k = \beta_k - 1;\ \beta^*{}_j = \nu - n + j,\ j = \overline{k+1,n-1};\ \alpha^*{}_i = N - n + i -$

$- \sum\limits_{m=1}^{j-1} \alpha^*{}_m, i = \overline{1,n},$ if $\alpha_i = 1, i = \overline{1,n}; \beta_k > \beta_{k-1} + 1;\ \beta_j = \beta_{j-1} + 1;\ j = \overline{k+1,n-1};$

3. $\beta^*{}_\mu = 0;\ \alpha^*{}_i = 0;\ V_{\beta^*{}_1 \ldots \beta^*{}_{n-1}; \alpha^*{}_1 \ldots \alpha^*{}_n}(\nu) = 0,$ if $\beta_\mu = \mu, \mu = \overline{1,n-1};\ \alpha_i = 1, i = \overline{1,n}.$

The expressions for the determination of the dispersion $D_{\alpha_1}(\nu)$, of casual coefficients $V_{\alpha_1}(\nu)$ are:

$$\begin{aligned} D_{\alpha_1}(\nu) = M\left[X^{2\alpha_1}(\nu)\right] - M^2\left[X^{\alpha_1}(\nu)\right] - \sum_{\lambda=1}^{\nu-1} \sum_{\xi_1^{(1)}=1}^{N-1} D_{\xi_1^{(1)}}(\lambda) \left\{ \varphi_{\xi_1^{(1)}}^{(\alpha_1)}(\lambda, \nu) \right\}^2 - \\ - \sum_{\xi_1^{(1)}=1}^{\alpha_1 - 1} D_{\xi_1^{(1)}}(\nu) \left\{ \varphi_{\xi_1^{(1)}}^{(\alpha_1)}(\nu, \nu) \right\}^2 - \\ - \sum_{\lambda=1}^{\nu-1} \sum_{l=2}^{M(\lambda)} \sum_{p_1^{(l)}=1}^{p_1'^{(l)}} \cdots \sum_{p_{l-1}^{(l)} = p_{l-2}^{(l)} + 1}^{p_{l-1}'^{(l)}} \sum_{\xi_1^{(l)}=1}^{\xi_1'^{(l)}} \cdots \sum_{\xi_l^{(l)}=1}^{\xi_l'^{(l)}} D_{p_1^{(l)} \ldots p_{l-1}^{(l)} \xi_1^{(l)} \ldots \xi_l^{(l)}}(\lambda) \times \\ \times \left\{ \varphi_{p_1^{(l)} \ldots p_{l-1}^{(l)} \xi_1^{(l)} \ldots \xi_l^{(l)}}^{(\alpha_1)}(\lambda, \nu) \right\}^2, \quad \nu = \overline{1,I}. \end{aligned} \tag{30}$$

Dispersions $D_{\beta_1,...\beta_{n-1};\alpha_1,...\alpha_n}(\nu)$ of casual coefficients $V_{\beta_1,...\beta_{n-1};\alpha_1,...\alpha_n}(\nu)$ are defined as

$$
\begin{aligned}
D_{\beta_1,...\beta_{n-1};\alpha_1,...\alpha_n}(\nu) &= M\left[X^{2\alpha_n}(\nu-\beta_{n-1})...X^{2\alpha_1}(\nu)\right]- \\
&-M^2\left[X^{\alpha_n}(\nu-\beta_{n-1})...X^{\alpha_1}(\nu)\right]-\sum_{\lambda=1}^{\nu-1}\sum_{\xi_1^{(1)}=1}^{N-1} D_{\xi_1^{(1)}}(\lambda)\times \\
&\times\left\{\varphi^{(\beta_1,...\beta_{n-1};\alpha_1,...\alpha_n)}_{\xi_1^{(1)}}(\lambda,\nu)\right\}^2- \\
&-\sum_{\lambda=1}^{\nu-1}\sum_{l=2}^{M(\lambda)}\sum_{p_1^{(l)}=1}^{p_1'^{(l)}}...\sum_{p_{l-1}^{(l)}=p_{l-2}^{(l)}+1}^{p_{l-1}'^{(l)}}\sum_{\xi_1^{(l)}=1}^{\xi_1'^{(l)}}...\sum_{\xi_l^{(l)}=1}^{\xi_l'^{(l)}} D_{p_1^{(l)}...p_{l-1;}^{(l)}\xi_1^{(l)}...\xi_l^{(l)}}(\lambda)\times \\
&\times\left\{\varphi^{(\beta_1,...\beta_{n-1};\alpha_1,...\alpha_n)}_{p_1^{(l)}...p_{l-1;}^{(l)}\xi_1^{(l)}...\xi_l^{(l)}}(\lambda,\nu)\right\}^2- \\
&-\sum_{l=2}^{n-1}\sum_{p_1^{(l)}=1}^{p_1'^{(l)}}...\sum_{p_{l-1}^{(l)}=p_{l-2}^{(l)}+1}^{p_{l-1}'^{(l)}}\sum_{\xi_1^{(l)}=1}^{\xi_1'^{(l)}}...\sum_{\xi_l^{(l)}=1}^{\xi_l'^{(l)}} D_{p_1^{(l)}...p_{l-1;}^{(l)}\xi_1^{(l)}...\xi_l^{(l)}}(\nu)\times \\
&\times\left\{\varphi^{(\beta_1,...\beta_{n-1};\alpha_1,...\alpha_n)}_{p_1^{(l)}...p_{l-1;}^{(l)}\xi_1^{(l)}...\xi_l^{(l)}}(\nu,\nu)\right\}^2- \\
&-\sum_{p_1^{(n)}=1}^{p_1^{*(n)}}...\sum_{p_{l-1}^{(n)}=p_{l-2}^{(n)}+1}^{p_{n-1}^{*(n)}}\sum_{\xi_1^{(n)}=1}^{\xi_1^{*(n)}}...\sum_{\xi_n^{(n)}=1}^{\xi_n^{*(n)}} D_{p_1^{(n)}...p_{n-1;}^{(n)}\xi_1^{(n)}...\xi_n^{(n)}}(\nu)\times \\
&\times\left\{\varphi^{(\beta_1,...\beta_{n-1};\alpha_1,...\alpha_n)}_{p_1^{(n)}...p_{n-1;}^{(n)}\xi_1^{(n)}...\xi_n^{(n)}}(\nu,\nu)\right\}^2,\nu=\overline{1,I}.
\end{aligned}
\tag{31}
$$

The coordinate functions of canonical decomposition (25) are defined by the formulas:

– to describe the relationship between the value $X^{\alpha_1}(\nu)$ and $X^{a_m}(i-b_{m-1})...X^{a_1}(i)$

$$\varphi_{\alpha_1}^{(b_1\ldots b_{m-1};a_1\ldots a_m)}(\nu,i)=\frac{1}{D_{\alpha_1}(\nu)}\Big\{M\Big[X^{\alpha_1}(\nu)X^{a_m}(i-b_{m-1})\ldots X^{a_1}(i)\Big]-$$

$$-M\Big[X^{\alpha_1}(\nu)\Big]M\Big[X^{a_m}(i-b_{m-1})\ldots X^{a_1}(i)\Big]-$$

$$-\sum_{\lambda=1}^{\nu-1}\sum_{\xi_1^{(1)}=1}^{N-1}D_{\xi_1^{(1)}}(\lambda)\varphi_{\xi_1^{(1)}}^{(\alpha_1)}(\lambda,\nu)\varphi_{\xi_1^{(1)}}^{(b_1\ldots b_{m-1};a_1\ldots a_m)}(\lambda,i)-$$

$$-\sum_{\xi_1^{(1)}=1}^{\alpha_1-1}D_{\xi_1^{(1)}}(\nu)\varphi_{\xi_1^{(1)}}^{(\alpha_1)}(\nu,\nu)\varphi_{\xi_1^{(1)}}^{(b_1\ldots b_{m-1};a_1\ldots a_m)}(\nu,i)- \tag{32}$$

$$-\sum_{\lambda=1}^{\nu-1}\sum_{l=2}^{M(\lambda)}\sum_{p_1^{(l)}=1}^{p_1'^{(l)}}\ldots\sum_{p_{l-1}^{(l)}=p_{l-2}^{(l)}+1}^{p_{l-1}'^{(l)}}\sum_{\xi_1^{(l)}}^{\xi_1'^{(l)}}\ldots\sum_{\xi_l^{(l)}}^{\xi_l'^{(l)}}D_{p_1^{(l)}\ldots p_{l-1;}^{(l)}\xi_1^{(l)}\ldots\xi_l^{(l)}}(\lambda)\times$$

$$\times\varphi_{p_1^{(l)}\ldots p_{l-1;}^{(l)}\xi_1^{(l)}\ldots\xi_l^{(l)}}^{(\alpha_1)}(\lambda,\nu)\varphi_{p_1^{(l)}\ldots p_{l-1;}^{(l)}\xi_1^{(l)}\ldots\xi_l^{(l)}}^{(b_1\ldots b_{m-1};a_1\ldots a_m)}(\lambda,\nu)\Big\},\ \nu=\overline{1,I}.$$

– to describe the relationship between the value $X^{\alpha_n}(\nu-\beta_{n-1})\ldots X^{\alpha_1}(\nu)$ and $X^{a_m}(i-b_{m-1})\ldots X^{a_1}(i)$

$$\varphi_{\beta_1,\ldots\beta_{n-1};\alpha_1,\ldots\alpha_n}^{(b_1,\ldots b_{m-1};a_1,\ldots a_m)}(\nu,i)=\frac{1}{D_{\beta_1,\ldots\beta_{n-1};\alpha_1,\ldots\alpha_n}(\nu)}\Big\{M\Big[X^{\alpha_n}(\nu-\beta_{n-1})\ldots X^{\alpha_1}(\nu)\times$$

$$\times X^{a_m}(i-b_{m-1})\ldots X^{a_1}(i)\Big]-M\Big[X^{\alpha_n}(\nu-\beta_{n-1})\ldots X^{\alpha_1}(\nu)\Big]\times$$

$$\times M\Big[X^{a_m}(i-b_{m-1})\ldots X^{a_1}(i)\Big]-M\Big[X^{\alpha_n}(\nu-\beta_{n-1})\ldots X^{\alpha_1}(\nu)\Big]\times \tag{33}$$

$$\times M\Big[X^{a_m}(i-b_{m-1})\ldots X^{a_1}(i)\Big]-$$

$$-\sum_{\lambda=1}^{\nu-1}\sum_{\xi_1^{(1)}=1}^{N-1}D_{\xi_1^{(1)}}^{(j)}(\lambda)\varphi_{\xi_1^{(1)}}^{(\beta_1,\ldots\beta_{n-1};\alpha_1,\ldots\alpha_n)}(\lambda,\nu)\varphi_{\xi_1^{(1)}}^{(b_1\ldots b_{m-1};a_1\ldots a_m)}(\lambda,i)-$$

$$-\sum_{\lambda=1}^{\nu-1}\sum_{l=2}^{M(\lambda)}\sum_{p_1^{(l)}=1}^{p_1'^{(l)}}\cdots\sum_{p_{l-1}^{(l)}=p_{l-2}^{(l)}+1}^{p_{l-1}'^{(l)}}\sum_{\xi_1^{(l)}=1}^{\xi_1'^{(l)}}\cdots\sum_{\xi_l^{(l)}=1}^{\xi_l'^{(l)}} D_{p_1^{(l)}\ldots p_{l-1;}^{(l)}\xi_1^{(l)}\ldots\xi_l^{(l)}}(\lambda)\times$$

$$\times\varphi^{(\beta_1,\ldots\beta_{n-1};\alpha_1,\ldots\alpha_n)}_{p_1^{(l)}\ldots p_{l-1;}^{(l)}\xi_1^{(l)}\ldots\xi_l^{(l)}}(\lambda,\nu)\varphi^{(b_1,\ldots b_{m-1};a_1,\ldots a_m)}_{p_1^{(l)}\ldots p_{l-1;}^{(l)}\xi_1^{(l)}\ldots\xi_l^{(l)}}(\lambda,i)-$$

$$-\sum_{l=2}^{n-1}\sum_{p_1^{(l)}=1}^{p_1'^{(l)}}\cdots\sum_{p_{l-1}^{(l)}=p_{l-2}^{(l)}+1}^{p_{l-1}'^{(l)}}\sum_{\xi_1^{(l)}=1}^{\xi_1'^{(l)}}\cdots\sum_{\xi_l^{(l)}=1}^{\xi_l'^{(l)}} D_{p_1^{(l)}\ldots p_{l-1;}^{(l)}\xi_1^{(l)}\ldots\xi_l^{(l)}}(\nu)\times$$

$$\times\varphi^{(\beta_1,\ldots\beta_{n-1};\alpha_1,\ldots\alpha_n)}_{p_1^{(l)}\ldots p_{l-1;}^{(l)}\xi_1^{(l)}\ldots\xi_l^{(l)}}(\nu,\nu)\varphi^{(b_1,\ldots b_{m-1};a_1,\ldots a_m)}_{p_1^{(l)}\ldots p_{l-1;}^{(l)}\xi_1^{(l)}\ldots\xi_l^{(l)}}(\nu,i)-$$

$$-\sum_{p_1^{(n)}=1}^{p_1^{*(n)}}\cdots\sum_{p_{l-1}^{(n)}=p_{l-2}^{(n)}+1}^{p_{n-1}^{*(n)}}\sum_{\xi_1^{(n)}=1}^{\xi_1^{*(n)}}\cdots\sum_{\xi_n^{(n)}=1}^{\xi_n^{*(n)}} D_{p_1^{(n)}\ldots p_{n-1;}^{(n)}\xi_1^{(n)}\ldots\xi_n^{(n)}}(\nu)\times$$

$$\times\varphi^{(\beta_1,\ldots\beta_{n-1};\alpha_1,\ldots\alpha_n)}_{p_1^{(n)}\ldots p_{n-1;}^{(n)}\xi_1^{(n)}\ldots\xi_n^{(n)}}(\nu,\nu)\varphi^{(b_1,\ldots b_{m-1};a_1,\ldots a_m)}_{p_1^{(n)}\ldots p_{n-1;}^{(n)}\xi_1^{(n)}\ldots\xi_n^{(n)}}(\nu,i)\Bigg\},\ \nu=\overline{1,I}.$$

Tipping in the expression (3) of known values $X(\mu)=x(\mu),\mu=\overline{1,k}$ converts the a priori random sequence into the a posteriori one:

$$X(i)=M\left[X(i)\right]+m_x^{(I-N+1,I-N+2,\ldots,I-1;1,1\ldots1,\nu)}(1,i)+$$

$$+\sum_{\nu=k+1}^{i-1}\sum_{\xi_1^{(1)}=1}^{N-1}V_{\xi_1^{(1)}}(\nu)\times\varphi^{(1)}_{\xi_1^{(1)}}(\nu,i)+V_1(i)+$$

$$+\sum_{\nu=k+1}^{i-1}\sum_{l=2}^{M(\nu)}\sum_{p_1^{(l)}=1}^{p_1'^{(l)}}\cdots\sum_{p_{l-1}^{(l)}=p_{l-2}^{(l)}+1}^{p_{l-1}'^{(l)}}\sum_{\xi_1^{(l)}=1}^{\xi_1'^{(l)}}\cdots\sum_{\xi_l^{(l)}=1}^{\xi_l'^{(l)}} V_{p_1^{(l)}\ldots p_{l-1;}^{(l)}\xi_1^{(l)}\ldots\xi_l^{(l)}}(\nu)\times \tag{34}$$

$$\times\varphi^{(1)}_{p_1^{(l)}\ldots p_{l-1;}^{(l)}\xi_1^{(l)}\ldots\xi_l^{(l)}}(\nu,i),\ i=\overline{k+1,I}.$$

Values of conditional expectation are defined as $(x(i)=m_x^{(\beta_1\ldots\beta_{n-1};\alpha_1\ldots\alpha_n,\nu)}(b_1\ldots b_{m-1};a_1\ldots a_m,i)$

$$x(i)=\begin{cases} M\left[X^{a_m}(i-b_{m-1})...X^{a_1}(i)\right], \ \nu=0; \\ m_x^{(\beta^*_1...\beta^*_{n-1};\alpha^*_1.....\alpha^*_n,\nu)}\left(b_1...b_{m-1};a_1...a_m,i\right)+\left[x^{\alpha_n}\left(\nu-\beta_{n-1}\right)...x^{\alpha_1}\left(\nu\right)\right. \\ \left.-m_x^{(\beta^*_1...\beta^*_{n-1};\alpha^*_1.....\alpha^*_n,\nu)}\left(\beta_1...\beta_{n-1};\alpha_1...\alpha_n,\nu\right)\right]\times \\ \times\varphi^{(b_1...b_{m-1};a_1...a_m)}_{\beta_1...\beta_{n-1};\alpha_1.....\alpha_n}\left(\nu,i\right),\text{if} \quad \alpha^*_1\neq 0,...,\alpha^*_n\neq 0; \\ m_x^{(p_1^{'(n-1)}...p_{n-2}^{'(n-1)};\xi_1^{'(n-1)}...\xi_{n-1}^{'(n-1)},\nu)}\left(b_1...b_{m-1};a_1...a_m,i\right)+\left[x^{\alpha_n}\left(\nu-\beta_{n-1}\right)...\right. \\ \left.\left.\left.\left.\left.\left.\left.\left.\left.\left....x^{\alpha_1}\left(\nu\right)-m_x^{(p_1^{'(n-1)}...p_{n-2}^{'(n-1)};\xi_1^{'(n-1)}...\xi_{n-1}^{'(n-1)},\nu)}\left(\beta_1...\beta_{n-1};\alpha_1...\alpha_n,\nu\right)\right]\times\right.\right.\right.\right.\right.\right.\right.\right.\right. \\ \times\varphi^{(b_1...b_{m-1};a_1...a_m)}_{\beta_1...\beta_{n-1};\alpha_1.....\alpha_n}\left(\nu,i\right), \text{ if } \ \alpha^*_1=0,...,\alpha^*_n=0. \end{cases} \tag{35}$$

Technology of predictive control of fail-safe operation on the model (35) is the same as using the expression (15).

4 Conclusion

Thus, there we obtained an calculation method for the estimation of the probability of fail-safe operation of information systems in the future instants of time. The technology is based on the model of the canonical expansion of the a posteriori random sequence of changes of the parameter controlled. Estimation of the probability of fail-safe operation of a information system based on the results of the numerical experiments is determined as a relative frequency of an event characterized by belonging of the realization to the allowable region on the forecast interval. The calculation method offered does not impose any significant constraints on the class of random sequences analyzed (linearity, stationarity, Markov behaviour, monotoneness, etc.). The only constraint is the finiteness of variance that is usually performed for real random processes. In contrast to the known solutions [12], [13] the suggested estimation procedure for fail-safe operation of information systems allows for nonlinear stochastic properties of the random sequence analyzed, which improves the accuracy of the predictive control procedure.

References

1. Frank, Paul M.: Fault Diagnosis in Dynamic Systems Using Analytical and Knowledge-Based Redundancy: a Survey and Some New Results, J. Automatica 3, 459-474 (1990)

2. Patton, Ron J., Paul M. Frank, Robert N. Clarke.: Fault Diagnosis in Dynamic Systems: Theory and Application, Prentice-Hall Inc. (1989)
3. Patton, Ron J., Robert N. Clark, Paul M. Frank: Issues of Fault Diagnosis for Dynamic Systems, Springer Science & Business Media (2000)
4. Frank, Paul M., Xianchun Ding: Survey of Robust Residual Generation and Evaluation Methods in Observer-Based Fault Detection Systems, Journal of Process Control 7(6), 403-424 (1997)
5. Ding S. X.: A Unified Approach to the Optimization of Fault Detection Systems, International Journal of Adaptive Control and Signal Processing 14(7), 725-745 (2000)
6. Silva N., Vieira M.: Towards Making Safety-Critical Systems Safer: Learning from Mistakes, In: ISSREW, 2014, 2014 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW), pp. 162-167 (2014)
7. Irrera, I., Duraes J., Vieira M.: On the Need for Training Failure Prediction Algorithms in Evolving Software Systems. In: High-Assurance Systems Engineering (HASE), 2014 IEEE 15th International Symposium on. IEEE, pp. 216-223 (2014)
8. Boyarchuk A.V., Brezhnev Ye.B., Gorbenko A.V., etc.: Safety of Critical Infrastructures: Mathematical Analysis and Engineering Methods of Analysis and Ensuring, National Aerospace University named after N.E.Zhukovsky "KhAI", Kharkiv (2011)
9. Palagin, A.V., Opanasenko, V.N.: Reconfigurable Computer Systems, Prosvita, Kyiv (2006)
10. Sokolov Y.N., Kharchenko V.S., Illyushko V.M., etc.: Applications of Computer Technologies for Software-Hardware Complexes Reliability and Safety Assessment Systems, National Aerospace University named after N.E.Zhukovsky "KhAI", Kharkiv (2013)
11. Odarushchenko O.N., Ponochovny Y.L., Kharchenko V.S., etc.: High Availability Systems and Technologies, National Aerospace University named after N.E.Zhukovsky "KhAI", Kharkiv (2013)
12. Kudritsky, V.D.: Predictive Control of Radioelectronic Devices, Technics, Kyiv (1982)
13. Kudritsky, V.D.: Filtering, Extrapolation and Recognition Realizations of Random Functions, FADA Ltd., Kyiv (2001)
14. Pugachev, V.S.: The Theory of Random Functions and its Application, Physmatgis, Moscow (1962)
15. Atamanyuk I.P., Kondratenko Y.P.: The Method of Forecasting Technical Condition of Objects. Patent 73885, Ukraine, G05B 23/02, Bul. № 19 (2012)
16. Atamanyuk, I.P., Kondratenko, V.Y., Kozlov, O.V., Kondratenko, Y.P.: The Algorithm of Optimal Polynomial Extrapolation of Random Processes, Modeling and Simulation in Engineering, Economics and Management, LNBIP 115, Springer, New York, 78-87 (2012)
17. Parzen, E.: On the Estimation of Probability Density Function and the Mode, J. Analysis of Mathematical Statistics. 33, 1065-1076 (1962)
18. Atamanyuk I.P.: Algorithm of Extrapolation of a Nonlinear Random Process on the Basis of Its Canonical Decomposition, J. Cybernetics and Systems Analysis. 41, Mathematics and Statistics, Kluwer Academic Publishers Hingham, 267-273 (2005)
19. Atamanyuk I.P., Kondratenko Y.P.: Algorithm of Extrapolation of a Nonlinear Casual Process on the Basis of Its Canonical Decomposition. In: The First International Workshop on Critical Infrastructure Safety and Security “CrISS-DESSERT–11”, Kharkiv, pp. 308-314 (2011)

Distributed Datastores: Towards Probabilistic Approach for Estimation of Reliability

Kyrylo Rukkas and Galyna Zholtkevych

V.N. Karazin Kharkiv National University
4, Svobody Sqr., 61022, Kharkiv, Ukraine
galynazholtkevych1991@gmail.com

Abstract. This paper focuses on the contradiction that follows from Brewer's Conjecture for distributed datastores: the need to deliver qualitative data to the user requires a guarantee of consistency, availability and stability of the system at the same time, but Brewer's Conjecture claims that this is impossible. To overcome this contradiction in the paper it is suggested to estimate statistically violation of these guarantees. To implement this idea the interdependencies between the guarantees and indicators of information quality are considered, different kinds of models specifying the general architecture and behaviour of datastores are described, and, finally, the basic metrics of guarantee ensuring are defined. This consideration allows us to formulate several problems that have both theoretical aspects and engineering applications for the improvement of the technology of distributed data processing.

Keywords: distributed datastore, CAP-theorem, information quality, statistic metrics, simulation
Key Terms: DistributedDataWarehousing, ConcurrentComputation

1 Introduction

Nowadays, any kind of human activity requires an infrastructure to support efficiently the corresponding process of decision making. A modern answer to such a requirement is an information system that is an integrated set of components to collect, store and process data, to deliver information, knowledge, and digital products. Today the development trend of Information and Communication Technology is a wide use of networking technologies. Therefore a typical modern information system is a distributed data processing system with a distributed datastore.

Now it is generally accepted that a distributed datastore should guarantee the following properties: consistency (C), availability (A), and partition tolerance (P). They are discussed in papers [5] and [1], but this discussion is too implicit. These works had been critically reviewed in [2]. This criticism is based on the fact that nobody had so far given explicit and rigorous definitions of these guarantees. Taking into account this remark we accept the following understanding as the origin point of our research:

consistency means that all replicas of any data unit contain the same data at the same time point;

availability means that a datastore eventually returns a response (either a success report or a failure notification) on each request;

and finally, ***partition tolerance*** characterizes the ability of a datastore to continue to operate despite arbitrary message losses or failure of part of the system (sometimes to refer to this ability the more general concept **fault-tolerance** is used).

In the ideal case a distributed datastore should provide these guarantees. In contrast to relational datastores like SQL databases that satisfy ACID properties and ensure the system safety, non-relational datastores do not provide complete safety for the information system. As known Brewer's conjecture [5], being called sometimes CAP theorem, says that it is impossible to maintain simultaneously all three CAP-guarantees in asynchronous or partially synchronous network and maintain safety in this way. Taking into account that synchronization of a distributed datastore decreases essentially system throughput we study consequences of Brewer's conjecture for asynchronous distributed datastores.

A lot of research had been wasted at the consideration of CAP-guarantees. As it is impossible to implement all three of them, we suggest a new approach: to characterise stochastically that each of these requirements is fulfilled or not (see Sec. 5) and do the best for a consumer - provide him with mechanisms used in different datastores for restoring the validity of the CAP-guarantees.

This supposition leads us to the need to develop a simulation framework for supporting numerical experiments to study these mechanisms.

As we said above, the principal objective for an information system design process is to provide a consumer with qualitative information just in time. Therefore we should understand how information quality (IQ) and the CAP-guarantees are connected. In order to clarify this interconnection we give some model of information quality and determine its indicators that depend on providing the CAP-guarantees in Sec. 2.

In Sec. 4 we present the conceptual model of a distributed datastore proposed as a background for the framework that should be developed. In Sec. 3 we discuss briefly the models for maintaining the consistency property in distributed systems. And finally, in Sec. 5 we give rigorous definitions for stochastic criteria of the CAP-guarantees providing, which is based on the presented conceptual model.

Further to avoid cumbersome formulations we shall say "CAP-properties" instead of "ensuring CAP-guarantees".

2 Interrelations between IQ model and CAP-properties

The problem to identify Information Quality (IQ) model was widely described and discussed in [3], [8], [4]. On our opinion, this discussion had been compactly summarized in [7]. The information represented in these sources correlates with

data quality standards in [6]. Below we give the list of IQ indicators according to this paper:

accessibility is the indicator that characterises the extent of availability and fast retrievability of information;

appropriate amount of information is used to denote if the scope of information is appropriate for the task at hand;

believability means that the information is considered as true and credible;

completeness evaluates whether information is sufficiently broad and deep to solve the task at hand;

concise representation characterises the compactness of data representation;

consistent representation means that all the information processes operate with the same data;

ease-of-manipulation envisages that information is easy manipulated and applied to different tasks;

free-of-error means that information is correct and reliable;

interpretability characterises that information is in appropriate languages, symbols and units, and the definitions are clear;

objectivity denotes that information is considered as unbiased and impartial;

relevancy envisages that information is applicable and helpful for the task at hand;

reputation determines that information should be highly regarded in the terms of its source or content;

security is satisfied if the access to information is restricted appropriately in accordance with the access rights;

timeliness is fulfilled if information is up-to-date for the task at hand;

understandability means that information is well-comprehended;

value-added assumes that information is beneficial and provides advantages from its use.

These indicators are used to define different views of an information system. We stress that the consideration is given from the next points of view: product and service quality. These quality indicators are also classified taking into account different views, namely, system specifications and consumer expectations. Therefore they had been grouped in a two-dimensional Table 1 presented below. (See more complete description of this classification in [7]).

Focusing on our subject, we consider only those of indicators, that depend on CAP-properties. Let us explain the reasons that has motivated us to take exactly these indicators of IQ model.

Obviously, by the definition the consistency has the direct impact on the consistent representation indicator. Also, consistency ensures the correct information, thus the free-of-error indicator directly depends on it. Furthermore, if the consistency is fulfilled, there may be a lot of replicas in the system, that is decreasing the concise representation indicator. And one can simply see that the consistency definition involves up-to-date information and then timeliness immediately depends on the consistency. Consistency does not have an influence

Table 1. Information Quality Model

	Conforms to Specifications	**Meets or Exceeds Consumer Expectations**
Product Quality	Sound Information: – Free-of-Error – Concise Representation – Completeness – Consistent representation	Useful Information – Appropriate Amount – Relevancy – Understandability – Interpretability – Objectivity
Service Quality	Dependable Information: – Timeliness – Security	Usable Information: – Believability – Accessibility – Ease-of-Manipulation – Reputation – Value Added

on the accessibility, because if the consistent information is received, it does not necessarily mean that it was retrieved quickly and easily.

Now we tell about the availability interrelations. Firstly, reasoning from the availability definition (see Sec. 1) the accessibility directly depends on the availability measurement. Secondly, the free-of-error indicator definition involves reliable data that is delivered just-in-time, so it also has a direct dependence on the availability. Evidently, it does not impact on the consistent representation and concise representation indicators; also it is not tightly connected with the timeliness indicator. Further in this paper we shall determine availability more strictly: we shall denote the availability as the meantime between a request and the response on it. And following from that definition, with improving availability the speed of retrieving data is increased, but it does not mean that these data are up-to-date.

The last group of interrelations is about the partition tolerance. It has an impact on the consistent representation, free-of-error, accessibility and timeliness quality indicators. The thing is that the perfect partition tolerance fulfillment ensures the successful response from a distributed datastore. In its turn, the successful response must always contain consistent, correct, reliable and up-to-date information. Otherwise the system answer is counted as an error message. Therefore greater probability to get the successful response gives higher values of indicators mentioned above.

Evidently, it does not have any influence on the concise representation in contrast to the consistency.

We summarize this connection in Table 2.

The endorsement of some interrelations and the discovering their behaviour can be obtained by the further experiments. The rest of indicators of IQ model

Table 2. Interrelations between IQ Indicators and CAP-requirements

	Consistency	**Availability**	**Partition Tolerance**
Consistent representation	+		+
Free-of-Error	+	+	+
Concise representation	+		
Accessibility		+	+
Timeliness	+		+

depend on the quality of the information system that provide information dataunits and we are not interested in them.

3 Brief overview of Models for Distributed Systems

The distributed datastores are various, therefore it is necessary to have tools for their analysis. The simplest classification is related to the ratio of read and write operations quantity. This classification should be reasonable from the point of view of three pillars that maintain all distributed systems: consistency, availability, partition tolerance. Hence the list will be as follows:

- Systems with domination of read operations (decision support or retrieval systems);
- Systems with domination of write operations (online transaction processing systems);
- Systems without explicit domination of read or write operations (business systems).

This classification is based on the quantity of read and write operations. It is important to know for us, because it may require different mechanisms for each type of system.

In this section we tell about the way to verify consistency and maintain the probability of its fulfillment following [10, Chapter 7]. An appropriate way to increase this probability is replication – making copies of new data units at each node and their updating.

Traditionally, consistency is discussed in the context of read and write operations in distributed systems. Following the book mentioned above there are two consistency models for different distributed systems: data-centric and client-centric ones. They are applied for different types of distributed systems. The first one, data-centric consistency model is a model for wide usage: online transaction and business systems mentioned above. It involves many types of consistency, such as continuous, sequential, causal and combined one, called grouping operations. The protocols for these consistency types are more complex than protocols in the second model. That is because data-centric model should be usable for systems where a lot of write operations may occur and spoil all the data units.

For instance, imagine if two employees in a company use the same datastore. They need to modify some file. And, unfortunately, it turns off that two employees download the file from the data storage and modify it in a different way and in different places. First one upload the file back and later the second one does the same. But the problem is that the file is modified in different places and there will be some conflicts. This is the simplest example, but if this datastore is distributed on many servers, there are a lot of copies of files in the data storage and more people use the same distributed system, it causes a complete mess in the storage. Therefore these protocols for controlling consistency should prevent such errors when a lot of write operations may occur.

The second one, client-centric model, is used for retrieval systems, where mostly read operations occur, but write ones are rare. Thus it is very costly and not necessary needed to use complex protocols. That is why for this model there exist such a type of the consistency as eventual consistency that ensures such a guarantee for the distributed system that eventually all the data units become same and consistency is fulfilled. The protocols for the client-centric model are also invented (see more in [10, Chapter 7]).

So as soon as we described different types of distributed datastores and established the problem to use different models and protocols for datastores that we focused on, we can go to the next section, where we specify the architecture and important behaviour elements for a distributed datastore.

4 Conceptual and Behavioral Model of Distributed Datastore

Above we described the general accepted model of the information quality that determines indicators which are needed for our research purpose, described models that are used for distributed datastore to satisfy consistency guarantees. But in order to come to the estimation step for the distributed datastore guarantees, obviously, we should also discuss the model of a distributed datastore.

Therefore in this section we represent the model of such systems in two views: structural and behavioural. Below there is given the structural one (Fig. 1).

The main component of our system is a distributed datastore. By definition it is a set of nodes (servers) connected by links between each other. Each node may have one or more neighbours. That is why this entity is composed of nodes and links. Obviously, each link is two-sided and a node entity may have many incidences.

Every node stores data units in replicas. If a node receives a new data unit it finds the data unit with the same identifier and replaces the old replica.

Nodes are classified into ready, busy and dead ones. If a node is busy or dead it ignores all the messages, so in this case the request will be lost. That is why the behaviour in this case is trivial.

The behaviour of a ready node is represented below, in the activity diagram (see Fig. 2).

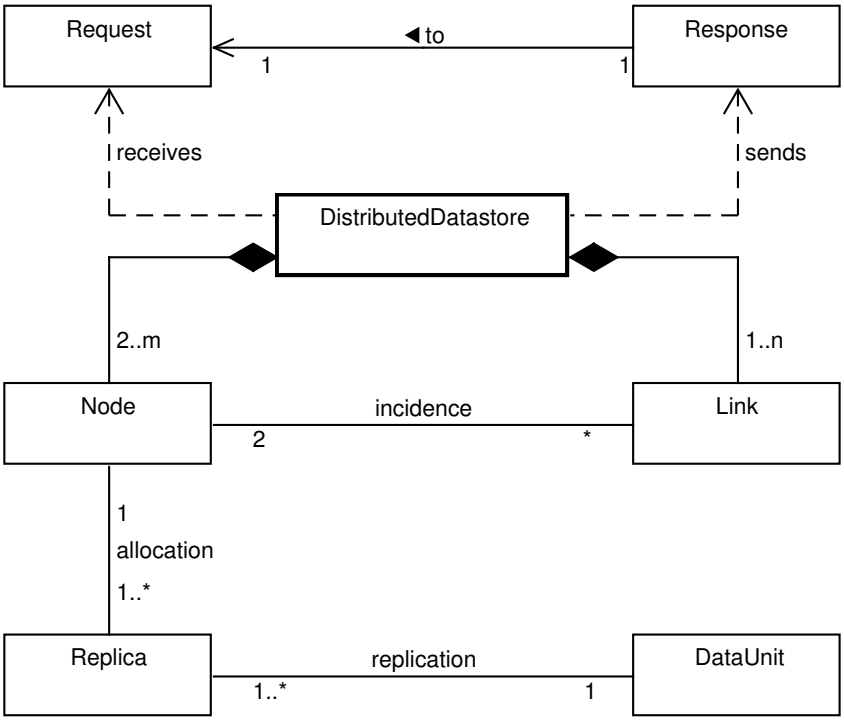


Fig. 1. Conceptual model

To present the behaviour in clear and understandable way, we separate more complex method of node *handle request* from the general behavioural view and show it in Fig. 3.

5 Distributed Datastores: Basic Prerequisites and Metrics

As said above the leading idea of this paper is to suggest an approach for estimating probabilistic metrics of CAP-properties.

It is generally accepted (see [9]) that a distributed datastore is a computer network where information is stored on more than one node, often in a replicated fashion.

Moreover, it is important to mention that a researcher has the possibility to observe datastore events in the sequence according to physical time while each

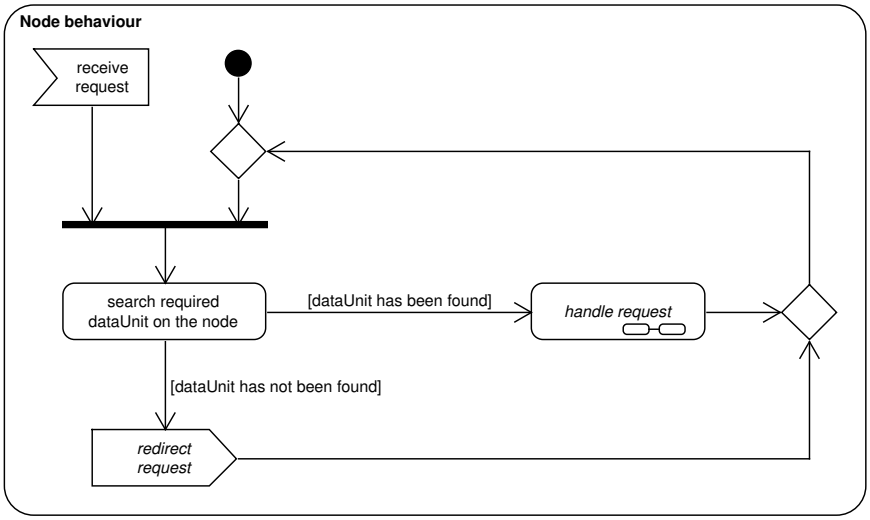


Fig. 2. Behaviour of node

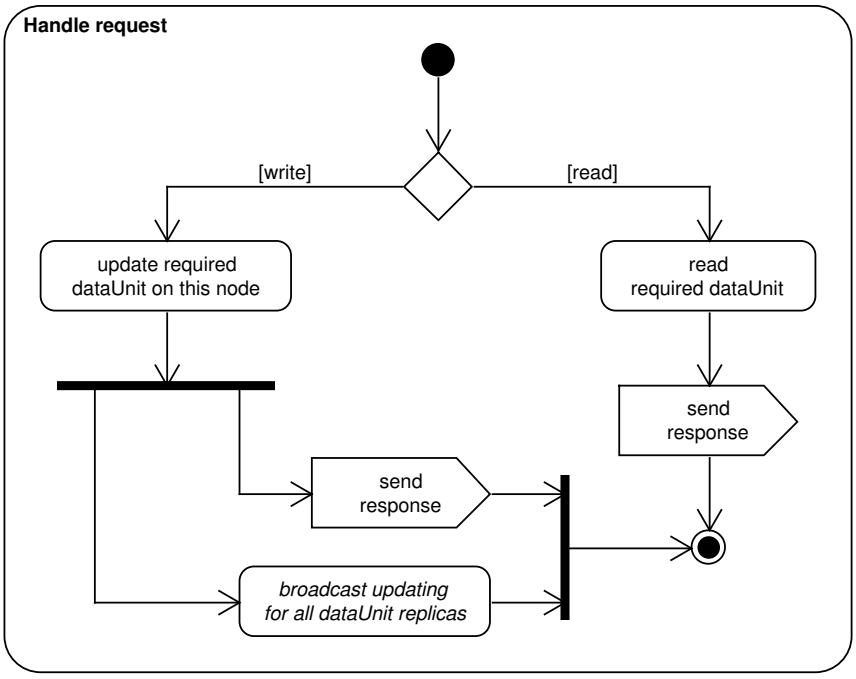


Fig. 3. Behaviour of handle request method

datastore node considers the sequence of events with respect to its local time only. Also we assume that datastores at study meet the eventual consistency requirement [10]. It means that after an isolated read request for any data unit all its replicas eventually have the same value.

Let us represent a short case study in order to understand the motivation to apply the probabilistic approach in asynchronous distributed datastore.

Let us suppose that there is the unreliable node in a distributed system that fails and recovers over over some time. At the initialization moment when nodes are established each node has the probabilistic distribution of recovering time and time of failure. In a distributed system links that bind nodes may also fail. To be able to calculate the probability of partition tolerance fulfillment we need to take into account these distributions in our research.

For now we are ready to describe how we apply the probabilistic approach in our studying by giving the rigorous definitions of CAP-properties. To do this, we describe a distributed datastore using the following mathematical model.

Our model is a tuple (N, L, ∂, D, r), where

N	is a finite set, whose elements correspond to nodes of a distributed datastore;
L	is a finite set, whose elements correspond to links of a distributed datastore;
$\partial : L \to 2^N$	is a mapping that associates each link with two nodes that it connects;
D	is a finite set, whose elements correspond to stored data units;
$r : D \to 2^N$	is a mapping that associates each data unit d with a subset of nodes that store its replica.

Thus, firstly, let us introduce the consistency metric. Taking into account that the consistency is the property when for each data unit its replicas have the same value, we shall consider the probability that all data units at distributed datastore are consistent at a certain time moment. Therefore we define the consistency metric in the following manner.

Definition 1 (Consistency metric). *Let $\sigma_t \in \{0, 1\}$ be the random variable that represents one of two events at time point $t \geq 0$:*

$\sigma_t = 0$	*corresponds to the event "there exists a data unit that is not consistent at the time point t" and*
$\sigma_t = 1$	*corresponds to the event "all data units are consistent at the time point t".*

Then the consistency metric $C(t)$ at time point t is defined by the formula

$$C(t) = \Pr(\sigma_t = 1)\,. \tag{1}$$

The second metric estimates the availability guarantee. We propose to measure the extent of availability as meantime between two events: the request has been received by the datastore and the corresponding response [1] has been sent by the datastore. More, formally:

[1] This response is either a successful report on request or an error message.

Definition 2. *Let τ_t be the time interval between a request receiving at the time point t and the corresponding response. Then the mean response time is defined by the formula*

$$T(t) = \mathbf{E}[\,\tau_t\,]\,. \tag{2}$$

Finally, to estimate the third guarantee, called the partition tolerance, we consider the ability of a datastore to survive network partitions, so that the performance of the datastore does not suffer. This definition is more complicated. Let us consider some time point t. At this instant some nodes are alive, but other ones failed. We denote by N_t^a the set of alive nodes ($N_t^a \subset N$). Similarly, we denote by L_t^a the set of alive links that connect alive nodes.

Definition 3. *We shall say that a data unit $d \in D$ is reachable from a node $n \in N_t^a$ at time point t if there exists a path in the graph (N_t^a, L_t^a) from n to some $n' \in N_t^a \bigcap r(d)$.*

Now we can introduce the metric for the partition tolerance using the previous definition.

Definition 4. *Let $\zeta_t \in \{0, 1\}$ be the random variable that represents one of two events at time point $t \geq 0$:*

- $\zeta_t = 0$ *corresponds to the event "there exists a data unit that is not reachable from some alive node at time point t" and*
- $\zeta_t = 1$ *corresponds to the event "all data units are reachable from any alive node at time point t".*

Then the partition tolerance metric $P(t)$ at time point t is defined by the formula

$$P(t) = \Pr(\zeta_t = 1)\,. \tag{3}$$

6 Conclusion

In this paper we have started studying the problem of the quality for distributed datastores. We have proposed the approach to measure the quality properties of a datastore. Therefore we have described CAP-properties and have built the metric system for CAP-properties estimation. We have described the indicators of the information quality and have found interrelations between the information quality and distributed datastores' properties basing on [7].

In order to have a view of the datastore and be able to work with it we have built its structural and behavioural model and based on this knowledge, we specified probabilistic characteristics for CAP-properties measurement.

Thus, the following steps for the problem set in the paper have been done:

- Formulation of understanding the idea: what are CAP-guarantees for distributed datastores indeed;
- Description of the information quality indicators;
- Investigating the connection between CAP-guarantees and information quality model;
- Building the structural and behavioural models for a distributed datastore;

– Forming "CAP-metrics" as the main idea for studying the quality of distributed datastores.

These steps give us possibilities to study CAP-properties fulfillment using the following background: the fault-tolerance mechanisms for asynchronous systems, concurrent programming, algorithms of data propagation in distributed systems, probably, some issues of internet strategy, Queueing Theory, Mathematical Statistics. Fault-tolerance protocols can be used to invent the algorithms for maintaining the partition tolerance guarantee. Obviously, a researcher should have the good background in Graph Theory, Probability Theory and Concurrent Programming in asynchronous distributed systems. Also, as we need to represent experimental results, it is necessary to imitate the model of a distributed datastore.

Summing up now we might set following problems:

– To simulate the model of a distributed datastore;
– To study different mechanisms to estimate the degree of ensuring each guarantee applying specified metrics;
– To analyse discovered mechanisms and their composition;
– To integrate these mechanisms with simulated asynchronous distributed datastore;
– To estimate theoretically the total time complexity for all provided mechanisms;
– To carry out the experimental estimation.

Acknowledgment. Authors express a deep gratefulness to Grygoriy Zholtkevych and Iryna Zaretska for their help and their criticism.

References

1. Brewer, E.: CAP Twelve Years Later: How the "Rules" Have Changed. Computer, IEEE Computer Society. Vol. 45, No. 2 (2012)
2. Burgess, M.: Deconstructing the 'CAP theorem' for CM and DevOps `http://markburgess.org/blog_cap.html` (2012)
3. CRG. Information Quality Survey: Administrators Guide. Cambridge Research Group, Cambridge, MA (1997)
4. English, L.P.: Information Quality Applied: Best Practices for Improving Business Information, Processes and Systems. Wiley (2009)
5. Gilbert, S., Lynch, N.: Brewers Conjecture and the Feasibility of Consistent, Available, Partition-Tolerant Web Services. ACM SIGACT News. Vol. 33, No. 2, pp. 51–59 (2002)
6. ISO 8000-1:2011. Data Quality. International Organisation for Standartization (2011).
7. Kahn, B.K., Strong, D.M., Wang, R.Y.: Information Quality Benchmarks: Product and Service Performance. Comm. ACM. Vol. 45, No. 4 (2002)
8. Kovac, R., Lee, Y.W., Pipino, L.L.: Total data quality management: the case of IRI. Conf. on Information Quality (Cambridge, MA), pp. 63–79 (1997)

9. Pessach, Ya.: Distributed Storage: Concepts, Algorithms, and Implementations. CreateSpace Independent Publishing Platform (2013)
10. Tanenbaum, A.S., Van Steen, M.: Distributed systems. Principles and Paradigms (2nd Edition). Prentice-Hall, Inc. (2006)

Direct Partial Logic Derivatives in Analysis of Boundary States of Multi-State System

Elena Zaitseva[1], Vitaly Levashenko[1], Jozef Kostolny[1], Miroslav Kvassay[1]

[1] University of Zilina, Department of Infromatics, Univerzitna 8215/1, 010 26, Zilina, Slovakia
{elena.zaitseva, vitaly.levashenko, jozef.kostolny, miroslav.kvassay}@fri.uniza.sk

Abstract. *Multi-State System* (MSS) is mathematical model that is used in reliability engineering for the representation of initial investigated object (system). In a MSS, both the system and its components may experience more than two states (performance levels). One of possible description of MSS is a structure function that is defined correlation between a system components states and system performance level. The investigation of a structure function allows obtaining different properties, measures and indices for MSS reliability. For example, boundary system's states, probabilities of a system performance levels and other measures are calculated based a structure function. In this paper mathematical approach of Direct Partial Logical Derivatives is proposed for calculation of boundary states of MSS.

Keywords. Multi-State System, Multiple-Valued Logic, Direct Partial Logic Derivatives, Boundary State

Key Terms. Reliability, Model, Approach, Methodology, ScientificField

1 Introduction

Reliability is a principal attribute for the operation of any modern technological system. A principal issue in reliability analysis is the uncertainty in the failure occurrences and consequences. With respect to the complexity of the system and the modeling of their behavior, a distinctive feature of the system reliability analysis is a comprehensive and integrated manner [1]. Focusing on safety, reliability engineering methods aim at the quantification of the probability of failure of the system. In paper [1] presents detail analysis of reliability engineering state and define principal problems in this scientific discipline. According to [1] one direction of reliability engineering is estimation of a complex system based on *Multi-State System* (MSS) reliability analysis methods.

MSS is mathematical model used in reliability analysis for system with some (more that two) levels of performance [1, 2]. This mathematical model has been exploited in reliability engineering since 1975 [3-5]. Principal advantage of this mathematical model is detail description of investigated object. MSS permits to

define and investigate several performance levels: from perfect function to fault. The typical Binary-State System allows evaluating only two system states as functioning and failure. Other states, for example, as partly functioning or functioning with restrictions are not analyzed in case of Binary-State System use. But extra states and performance levels in the mathematical model dramatically increase size of this model and computational complexity of its analysis. Therefore the MSS has not been used intensively in reliability analysis. There is other aspect to restrict the application of MSS. It is absent of effective methods for MSS analysis.

According to analysis in paper [2] there are four groups of methods for MSS analysis that are based on different mathematical approach: an extension of Boolean models to the multi-valued case, the stochastic process as Markov process, the universal generation function methods and the Monte-Carlo simulation techniques. For example, Markov processes are used to analyze the system state transition process [6]. The universal generation function application is useful in optimization problem [7]. The Monte-Carlo simulation as a rule is used for reliability assessment of system with large number of components [8]. The methods based on extension of Boolean models to the multi-valued case were developed historically the first [9, 10]. According to these methods MSS is represented and defined by the structure function. This function is defined the conformance of the system performance level and components states. As a rule for the structure function definition and representation is used Boolean functions [10, 11]. Only in separated publications the structure function with some values has been considered [9, 12]. In papers [13, 14] the correlation of *Miltiple-Valued Logic* (MVL) function and structure function was analyzed. The interpretation of the structure function as the MVL function allowed using the mathematical approach of MVL in the analysis of the MSS structure function. In paper [14] the application of Logical Differential Calculus for MSS reliability analysis has been proposed. The Logic Differential Calculus is used for analysis of dynamic properties of MVL function and this approach can be applied for analysis of dynamic behaviour of MSS that is determined by the structure function.

In the paper [14] the basic conception of application of Direct Partial Logic Derivatives (it is part of Logical Differential Calculus) in MSS reliability analysis has been considered. The proposed method permits to investigate the influence of one system component state change to the system performance level. The new indices for quantitative analysis of such influence have been defined. The application of these derivatives for calculation of importance measures (Structural Importance, Birnbaum Importance and Criticality Importance) has been investigated in paper [15, 16]. The algorithm for calculation of Fussell-Vesely Importance based on Direct Partial Logic Derivatives has been proposed in [17].

In this paper new application of Direct Partial Logic Derivatives for MSS reliability analysis based on the structure function is considered. The new algorithm for calculation of boundary states of MSS is proposed. This algorithm is based on representation of MSS by the structure function (section 2). This section includes the conception of boundary states of MSS for every performance level. The special structures of MSS (parallel and series) and their structure function are considered in the section 2 too. These types of system are typically employed in reliability analysis [18]. The short description of conceptions of Direct Partial Logic Derivatives with respect to one variable and with respect to variable vector are presented in the section

3. The calculation of MSS boundary states for every system performance level is considered section 4.

2 MSS Structure Function

2.1 Structure Function of MSS

Consider the system of n components. Each component of this system has m states: from the complete failure (it is 0) to the perfect functioning (it is m-1). The i-th system component state is denoted as x_i (i = 1, …, n). Suppose, that this system has m performance level too: from the complete failure (it is 0) to the perfect functioning (it is m-1). The dependence of the system performance level on components states is defined by the structure function $\phi(\boldsymbol{x})$ identically:

$$\phi(x_1, x_2,\ldots, x_n) = \phi(\boldsymbol{x})\text{: } \{0, 1, \ldots, m\text{-}1\}^n \rightarrow \{0, 1, \ldots, m\text{-}1\}. \quad (1)$$

The function (1) agrees with the definition of a MVL function [19]. Therefore the mathematical approaches of MVL can be used in quantification analysis of MSS. But the structure function (1) allows representing the very small class of real system for which the number of system performance levels and number of every component states are equal. As a rule, the real-world system has different numbers of states for different components. And the number of performance levels can be different too. Therefore the structure function of real-word system must be defined as:

$$\phi(\boldsymbol{x})\text{: } \{0, 1, \ldots, m_1\text{-}1\}\times\ldots\times\{0, 1, \ldots, m_n\text{-}1\} \rightarrow \{0, 1, \ldots, M\text{-}1\}, \quad (2)$$

where m_i is number of states for i-th system component (i = 1, …, n) and M is number of a system performance levels.

The equation (2) can be interpreted as a MVL function. But some formal changes allow transforming this structure function definition into an incompletely specified MVL function. This transformation suppose the interpretation of the function (2) as incompletely specified MVL function for maximal value of number m_i and M: $m_{\max}$=MAX$\{m_1, \ldots, m_n, M\}$. In this case, the structure function (2) is defined as:

$$\phi`(\boldsymbol{x})\text{: } \{0, 1, \ldots, m_{\max}\text{-}1\}^n \rightarrow \{0, 1, \ldots, m_{\max}\text{ -}1\}. \quad (3)$$

The interpretation of the structure function (2) as an incompletely specified MVL function (3) permits to use mathematical approaches of MVL without principal restriction for analysis of properties of the structure function (2).

For example, consider the simple service system (Fig. 1) in a region from paper [17]. This system consists of three components (n = 3) – service point 1 (x_1), service point 2 (x_2) and infrastructure (x_3). This system has four performance levels: 0 – non-operational (no customer is satisfied), 1 – partially operational (some customers are satisfied), 2 – partially non-operational (some customers are not satisfied), 3 – fully

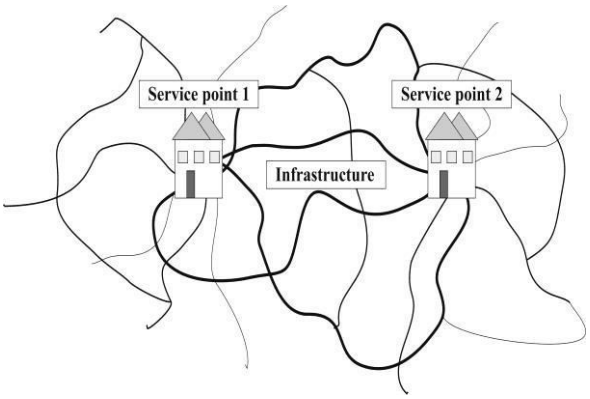


Fig. 1. A simple service system.

Table 1. The structure function of the simple service system

Components states		x_3			
x_1	x_2	0	1	2	3
0	0	**0**	**0**	**0**	**0**
0	1	**0**	**1**	**1**	**2**
1	0	**0**	**1**	**1**	**2**
1	1	**0**	**2**	**3**	**3**

Table 2. The structure function of the simple service system represented as an incompletely specified MVL function

Components states		x_3			
x_1	x_2	0	1	2	3
0	0	**0**	**0**	**0**	**0**
0	1	**0**	**1**	**1**	**2**
0	2	–	–	–	–
0	3	–	–	–	–
1	0	**0**	**1**	**1**	**2**
1	1	**0**	**2**	**3**	**3**
1	2	–	–	–	–
1	3	–	–	–	–
2	0	–	–	–	–
2	1	–	–	–	–

2	2	–	–	–	–
2	3	–	–	–	–
3	0	–	–	–	–
3	1	–	–	–	–
3	2	–	–	–	–
3	3	–	–	–	–

operational (all customers are satisfied). Next, we assume that the service points are only functional (state 1) or dysfunctional (state 0). The infrastructure can be modelled by 4 quality levels, i.e. from 0 (the quality of the infrastructure is poor) to 3 (the quality is perfect). The structure function of this system according to (2) is defined in Table 1 ($m_1 = m_2 = 2$, $m_3 = 4$ and $M = 4$). The structure function of this system as an incompletely specified MVL function is shown in Table 2 ($m_{\max} = 4$).

2.2 Series and Parallel MSS

There are some typical structures in the reliability engineering: series, parallel, *k*-out-of-*n* and bridge. Every system of these structures has single valued definition for Binary-State System (Fig.2).

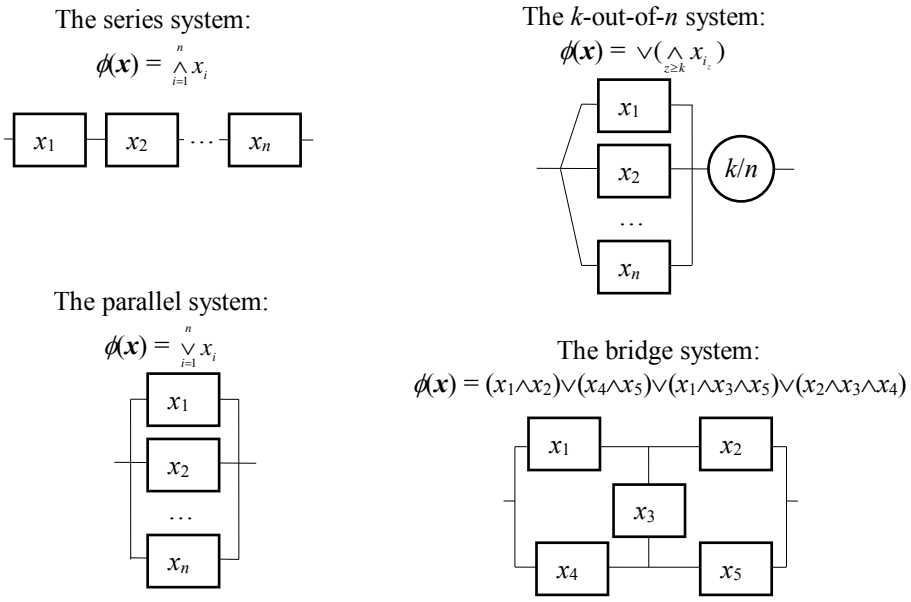


Fig. 2. Graphical and mathematical interpretation of typical structures of Binary-State System.

The extension of mathematical definition of the structure functions of these system for the MSS has some variants. For example, in paper [24] the structure functions of series, parallel and *k*-out-of-*n* MSS are defined based on the interpretation of mathematical equations of these system in terms of MVL. Structure

functions of these MSS in MVL terms are declared by OR ($\vee$) and AND ($\wedge$) MVL functions: OR(a, b) = MAX(a, b) and AND(a, b) = MIN(a, b). According to [24] structure functions parallel, series and k-out-of-n MSS are declared

- for parallel MSS as:

$$\phi(\boldsymbol{x}) = \bigvee_{i=1}^{n} x_i = \mathrm{MAX}(x_1, x_2, \ldots, x_n), \tag{4}$$

- for series MSS as:

$$\phi(\boldsymbol{x}) = \bigwedge_{i=1}^{n} x_i = \mathrm{MIN}(x_1, x_2, \ldots, x_n), \tag{5}$$

- for k-out-of-n MSS as:

$$\phi(\boldsymbol{x}) = \vee(\underset{z \geq k}{\wedge} x_{i_z}) = \mathrm{MAX}(\mathrm{MIN}_1(x_{i_1}, x_{i_2}, \ldots, x_{i_k}), \ldots, \mathrm{MIN}_w(x_{i_1}, x_{i_2}, \ldots, x_{i_k})), \tag{6}$$

$w = n!/(k!(n-k)!)$.

In paper [27] other declaration of these MSS are presented in which a system performance level depends on the number of functioning components. In Table 3 some structure functions of parallel MSS of two components (n = 2) with three states (m_1 = m_2 =3) and three performance level (M = 3) are shown. All these structure function in case of Binary-State System are parallel system. The series MSS can be defined similarly.

Table 3. The structure functions of parallel MSS (n = 2, m_1 = m_2 = M = 3)

Components states		Structure function of parallel MSS			
x_1	x_2	$\phi_1(\boldsymbol{x})$	$\phi_2(\boldsymbol{x})$	$\phi_3(\boldsymbol{x})$	$\phi_4(\boldsymbol{x})$
0	0	**0**	**0**	**0**	**0**
0	1	**1**	**1**	**1**	**1**
0	2	**2**	**1**	**1**	**1**
1	0	**1**	**1**	**1**	**1**
1	1	**1**	**2**	**1**	**1**
1	2	**2**	**2**	**1**	**2**
2	0	**2**	**1**	**1**	**1**
2	1	**2**	**2**	**1**	**2**
2	2	**2**	**2**	**2**	**2**

Therefore the typical structures of MSS have no single valued mathematical definition, because there are the set of structure functions of MSS that can be agreed with one Binary-State System. The structure function allows defining MSS explicitly. Therefore the structure function is preferable form of MSS mathematical representation.

2.3 Boundary States of MSS

The conception of boundary states has been proposed for Binary-State System firstly. The boundary state is defined as state for which the failure of one system components or some components causes the fault of a system [20]. The boundary state of MSS must be defined for every system performance level [2]. In papers [21, 22] the boundary states of MSS are interpreted as minimal cut/path sets. Authors of [23] introduced conception of Lower (Upper) Boundary Points of MSS for system performance level j (j =0, …, M-1). The boundary states for system performance level j and component i (i = 1, …, n) (named as exact boundary states) has been proposed and considered in papers [24, 25]. In paper [26] and [17] the correlations of these boundary states with Lower (Upper) Boundary Points and minimal cut/path sets are shown accordingly.

The exact boundary states have been considered in paper [25]. These states are system states for which the change of the i-th component state from s to $\tilde{s}$ causes the system performance level change from j to $\tilde{j}$ (s, $\tilde{s} \in \{0,\ldots, m_i\text{-}1\}$, $s \neq \tilde{s}$ and j, $\tilde{j} \in \{0,\ldots, M\text{-}1\}$, $j \neq \tilde{j}$). The exact boundary state is defined by the exact boundary vector unambiguously. Illustrate the correlation of a system exact boundary state and an exact boundary vector by the example for the service system in Fig.1.

Determine the exact boundary states of this service system for which the failure of the first component causes the system failure as the change of the system performance level from state “1” to “0”. According to Table 1, there are two situations that agree to this condition. They are possible for the failure of the second component and the third component state “1”or “2”. These exact boundary states can be presented as vector states: $\boldsymbol{x} = (x_1, x_2, x_3) = (1 \rightarrow 0,0,1)$ and $\boldsymbol{x} = (x_1, x_2, x_3) = (1 \rightarrow 0,0,2)$. Note that the boundary state $\boldsymbol{x} = (x_1, x_2, x_3) = (1 \rightarrow 0,0,3)$ does not satisfy the condition because the system performance level changes from “1” to “2” depending on the failure of the first component in this case.

One of possible mathematical approaches for the definition of the exact boundary states in MVL is Logical Differential Calculus, in particular the Direct Partial Logical Derivatives. Consider the application of this mathematical approach for analysis of structure function of MSS.

3 Direct Partial Logical Derivatives

3.1 Direct Partial Logical Derivative with respect to one variable

The mathematical tool of Direct Partial Logic Derivatives has been proposed in [25] for calculation of an exact boundary states of a MSS. The Direct Partial Logic Derivative with respect to variable x_i for the structure function (1) permits to analyze the system performance level change from j to $\tilde{j}$ when the i-th component state changes from s to $\tilde{s}$:

$$\partial\phi(j \to \tilde{j})/\partial x_i(s \to \tilde{s}) = \begin{cases} 1, \text{ if } \phi(s_i, \boldsymbol{x}) = j \text{ and } \phi(\tilde{s}_i, \boldsymbol{x}) = \tilde{j} \\ 0, \text{ other} \end{cases}, \quad (7)$$

where $\phi(s_i, \boldsymbol{x}) = \phi(x_1,\ldots, x_{i-1}, s, x_{i+1},\ldots, x_n)$; $\phi(\tilde{s}, \boldsymbol{x}) = \phi(x_1,\ldots, x_{i-1}, \tilde{s}, x_{i+1},\ldots, x_n)$; $s, \tilde{s} \in \{0,\ldots, m_i - 1\}, s \neq \tilde{s}$ and $j, \tilde{j} \in \{0,\ldots, M - 1\}, j \neq \tilde{j}$.

For example, investigate the influence of the first component failure to the fault of the simple service system in Fig.1. The Direct Partial Logic Derivative $\partial\phi(1\to0)/\partial x_1(1\to0)$ allows to calculate the system state for which this failure causes the system break down. The calculation of this derivative is shown in Fig.3 in form of flow graph. The derivative $\partial\phi(1\to0)/\partial x_1(1\to0)$ has two non-zero values that agrees with state vectors: $\boldsymbol{x} = (x_1, x_2, x_3) = (1 \to 0,0,1)$ and $\boldsymbol{x} = (x_1, x_2, x_3) = (1 \to 0,0,2)$. According to definition of the Direct Partial Logic Derivative (7) for these system state the failure of the first system component causes the system failure too. Therefore the service system fails after the failure of the first service point if the second service point isn't functioning and the functioning of the infrastructure conforms to state one or two. The system states $\boldsymbol{x} = (x_1, x_2, x_3) = (1 \to 0,0,1)$ and $\boldsymbol{x} = (x_1, x_2, x_3) = (1 \to 0,0,2)$ are exact boundary states for the first system component failure and the system performance level 1.

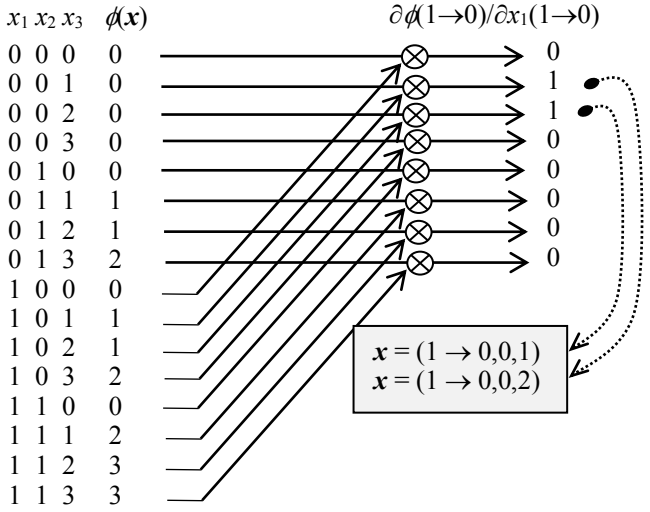


Fig. 3. Calculation of the Direct Partial Logic Derivative $\partial\phi(1\to0)/\partial x_1(1\to0)$.

The Direct Partial Logic Derivative (7) allows investigating boundary states of a MSS for which component state x_i change from s to $\tilde{s}$ causes the system performance level change from j to $\tilde{j}$. Therefore, this derivative allows calculating exact boundary states of the i-th system component for MSS performance level j that agree to state vectors $\boldsymbol{x} = (x_1, x_2,\ldots, x_n)$. All possible changes of the i-th system

component and their influence to MSS performance level can be investigated based on the Direct Partial Logic Derivative (7). But this derivative permits to investigate the influence of one component only. The Direct Partial Logic Derivative with respect of variable vector investigate the system state changes depending on changes of states of some system components.

3.2 Direct Partial Logical Derivative with respect to variable vector

A Direct Partial Logic Derivatives of a structure function $\phi(\boldsymbol{x})$ of n variables with respect to variables vector $\boldsymbol{x}^{(p)} = (x_{i_1}, x_{i_2}, \ldots, x_{i_p})$ reflects the fact of changing of function from j to $\tilde{j}$ when the value of every variable of vector $\boldsymbol{x}^{(p)}$ is changing from s to $\tilde{s}$ [15]:

$$\partial\phi(j \to \tilde{j})/\partial x_i(\boldsymbol{s}^{(p)} \to \tilde{\boldsymbol{s}}^{(p)}) = \begin{cases} 1, \text{ if } \phi(s_{i_1},\ldots,s_{i_p},\boldsymbol{x}) = j \text{ and } \phi(\tilde{s}_{i_1},\ldots,\tilde{s}_{i_p},\boldsymbol{x}) = \tilde{j} \\ 0, \text{ other} \end{cases} . \quad (8)$$

In (8) a change of value of i_q-th variable x_{i_q} form s_{i_q} to $\tilde{s}_{i_q}$ agrees with a change of i_q-th MSS component state form s_{i_q} to $\tilde{s}_{i_q}$ (q = 1, …, p and $p < n$). So, changes of some components states correspond with change of a variables vector $\boldsymbol{x}^{(p)} = (x_{i_1}, x_{i_2}, \ldots, x_{i_p})$. Every variable values of this vector changes form s_{i_q} to $\tilde{s}_{i_q}$. So, vector $\boldsymbol{x}^{(p)}$ can be interpreted as components states vector or components efficiencies vector.

For example, consider the simple service system (Fig.1) failure depending on fault of the first service point and reduction of functioning of infrastructure from state 2 to 1. This system behavior can be presented by the Direct Partial Logic Derivative $\partial\phi(1\to0)/\partial x_1(1\to0)\ \partial x_3(2\to1)$. The calculation of this derivative is shown in Fig.4 by the flow graph.

The derivative $\partial\phi(1\to0)/\partial x_1(1\to0)\ \partial x_3(2\to1)$ has two values and one of them is non-zero value that agrees with state vector: $\boldsymbol{x} = (x_1, x_2, x_3) = (1 \to 0, 0, 2 \to 1)$. This state vector define of the service system failure depending on the failure of the first service point and deterioration of the infrastructure functioning from state 2 to state 1. Therefore the system state $\boldsymbol{x} = (x_1, x_2, x_3) = (1 \to 0, 0, 2 \to 1)$ can be interpreted as exact boundary state for the first and the third system components of the system performance level 1.

The Direct Partial Logic Derivative with respect to variable vector (8) allows investigating boundary states of a MSS for which simultaneous changes of p components states from s_{i_q} to $\tilde{s}_{i_q}$ (q = 1, …, p and $p < n$) causes the system performance level change from j to $\tilde{j}$. Therefore, the Direct Partial Logic Derivative with respect to variable vector allows calculating exact boundary states for MSS performance level j of the i-th system component.

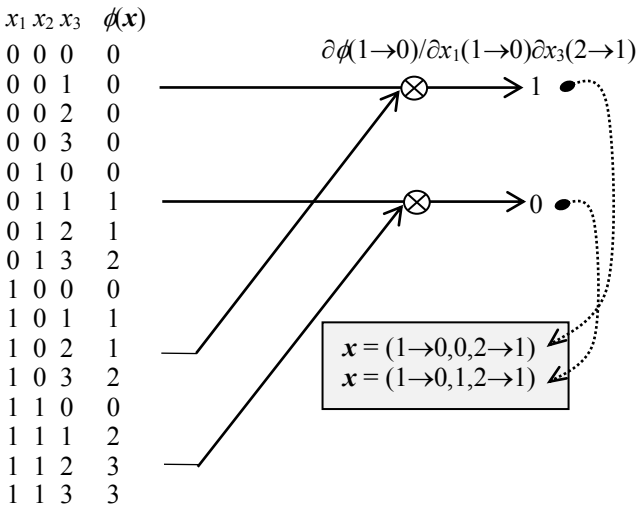


Fig. 4. Calculation of the Direct Partial Logic Derivative $\partial\phi(1\to0)/\partial x_1(1\to0)\,\partial x_3(2\to1)$.

4 The Calculation and Estimation of Exact Boundary States of MSS based on Direct Partial Logic Derivatives

The exact boundary state of MSS are defined based on the condition that fixed system performance level change depending on the appointed change of one system component state or specified changes of some components states. The Direct Partial Logic Derivative with respect to one variable (7) and the Direct Partial Logic Derivative with respect to variable vector (8) can be used to investigate change of the system performance level from j to $\tilde{j}$ that are caused by specified changes of one or some system components states. These derivatives have non-zero values of the structure function for system states that satisfy for specified condition: the system performance level change from j to $\tilde{j}$ depending on specified changes of one or some system components states. *Therefor the exact boundary states can be defined as system states that conform to non-zero values of derivatives (7) and (8).*

The exact boundary state for MSS performance level j depending on the i-th system component $\begin{pmatrix} j\to\tilde{j} \\ x_i \\ s\to\tilde{s} \end{pmatrix}$ is indicated by vector state $\boldsymbol{x} = (x_1,\ldots, x_i,\ldots, x_n) = (a_1,\ldots, s_i,\ldots, a_n)$ for which $\phi(a_1,\ldots, s_i,\ldots, a_n) = j$ and $\phi(a_1,\ldots, \tilde{s}_i,\ldots, a_n) = \tilde{j}$. This state is calculate as non-zero value of Direct Partial Logic Derivative (7). The exact boundary state for MSS performance level j depending on p components $x_{i_1}, x_{i_2}, \ldots,$

x_{i_p} $\begin{pmatrix} {\scriptstyle j\to\tilde{j}} \\ x_{i_1}\ldots x_{i_p} \\ {\scriptstyle s_{i_1}\to\tilde{s}_{i_1} s_{i_p}\to\tilde{s}_{i_p}} \end{pmatrix}$ is indicated by vector state $\boldsymbol{x} = (x_1,\ldots, x_{i_1},\ldots, x_{i_p},\ldots, x_n) = (a_1,\ldots, s_{i_1},\ldots, s_{i_p},\ldots, a_n)$ for which $\phi(a_1,\ldots, s_{i_1},\ldots, s_{i_p},\ldots, a_n) = j$ and $\phi(a_1,\ldots,\tilde{s}_{i_1},\ldots,\tilde{s}_{i_p},\ldots,a_n) = \tilde{j}$. This state is calculate as non-zero value of Direct Partial Logic Derivative (8).

Consider estimation of the system boundary states for coherent MSS. There are next assumptions for structure function of coherent MSS [2]: (*a*) the structure function is monotone and $\phi(\boldsymbol{s})=s$ ($s\in\{0, \ldots, m\text{-}1\}$) and (*b*) all components are *s*-independent and are relevant to the system.

Every system component is characterized by the probabilities of its state:

$$p_{i,s} = \Pr\{x_i = s\},\ s = 0, \ldots, m_i\text{-}1. \tag{9}$$

The probability of every boundary state (a_1, …, s_i, …, a_n) for MSS performance level *j* depending on the *i*-th system component change from *s* to $\tilde{s}$ is calculated based on the probabilities of components states:

$$p_{(a_1\ldots s_i\ldots a_n)}\begin{pmatrix} {\scriptstyle j\to\tilde{j}} \\ x_i \\ {\scriptstyle s\to\tilde{s}} \end{pmatrix} = p_{1,a_1}\cdot\ldots\cdot p_{i-1,a_{i-1}}\cdot p_{i,s_i}\cdot p_{i+1,a_{i+1}}\cdot\ldots p_{n,a_n}. \tag{10}$$

The probability of boundary state for MSS performance level *j* depending on the *i*-th system component states changes is calculated as:

$$p\begin{pmatrix} {\scriptstyle j\to\tilde{j}} \\ x_i \\ {\scriptstyle s\to\tilde{s}} \end{pmatrix} = \sum_{s,\tilde{s}} p_{(a_1\ldots s_i\ldots a_n)}\begin{pmatrix} {\scriptstyle j\to\tilde{j}} \\ x_i \\ {\scriptstyle s\to\tilde{s}} \end{pmatrix}. \tag{11}$$

Next measure is defined the probability of boundary state of MSS performance level *j* depending of all component state change from *s* to $\tilde{s}$:

$$p\begin{pmatrix} {\scriptstyle j\to\tilde{j}} \\ x \\ {\scriptstyle s\to\tilde{s}} \end{pmatrix} = \sum_{i=1}^{n} p\begin{pmatrix} {\scriptstyle j\to\tilde{j}} \\ x_i \\ {\scriptstyle s\to\tilde{s}} \end{pmatrix}. \tag{12}$$

The probability of MSS performance level change depending on the change of the *i*-th system component from state *s* to $\tilde{s}$ is calculated according to:

$$p\left(\begin{matrix} x_i \\ s\to\tilde{s} \end{matrix}\right)=\sum_{j,\tilde{j}} p\left(\begin{matrix} x_i^{j\to\tilde{j}} \\ s\to\tilde{s} \end{matrix}\right). \quad (13)$$

The probability of MSS performance level change depending on all change of the i-th system component state is generalization of previous equation:

$$p\left(x_i\right)=\sum_{s,\tilde{s}} p\left(\begin{matrix} x_i \\ s\to\tilde{s} \end{matrix}\right). \quad (14)$$

The similar measures to (10) - (14) can be defined for estimation of exact boundary state for MSS performance level j of p components $x_{i_1}, x_{i_2}, \ldots, x_{i_p}$ $\left(\begin{matrix} x_{i_1}^{j\to\tilde{j}} \ldots x_{i_p} \\ s_{i_1}\to\tilde{s}_{i_1} s_{i_p}\to\tilde{s}_{i_p} \end{matrix}\right)$.

Consider some examples for calculation of measures (10) - (14) for the simple service system in Fig.1. The components states probabilities for this system are defined in Table 4. Consider this system failure depending to the first components. The Direct Partial Logic Derivative $\partial\phi(1\to0)/\partial x_1(1\to0)$ represents this system behavior (Fig.3). This derivative has two non-zero values that conform to two boundary states $\left(\begin{matrix} x_1^{1\to0} \\ 1\to0 \end{matrix}\right)$: $\boldsymbol{x}=(1,0,1)$ and $\boldsymbol{x}=(1,0,2)$.

Table 4. The components states probabilities of the simple service system

0	Components states							
	x_1		x_2		x_3			
	0	1	0	1	0	1	2	3
$p_{i,s}$	0.3	0.7	0.2	0.8	0.2	0.6	0.1	0.1

The probabilities of boundary states for the system failure depending the first component break down $\left(\begin{matrix} x_1^{1\to0} \\ 1\to0 \end{matrix}\right)$ are calculate according to (10) and are:

$$p_{(101)}\left(\begin{matrix} x_1^{1\to0} \\ 1\to0 \end{matrix}\right)=p_{1,1}\cdot p_{2,0}\cdot p_{3,1}=0.084 \text{ and } p_{(102)}\left(\begin{matrix} x_1^{1\to0} \\ 1\to0 \end{matrix}\right)=p_{1,1}\cdot p_{2,0}\cdot p_{3,2}=0.014 \quad (15)$$

The probability of boundary state for this system failure depending on the first component is calculated based on (11) as:

$$p\begin{pmatrix}1\to 0\\ x_1\\ 1\to 0\end{pmatrix} = p_{(10)}\begin{pmatrix}1\to 0\\ x_1\\ 1\to 0\end{pmatrix} + p_{(10)}\begin{pmatrix}1\to 0\\ x_1\\ 1\to 0\end{pmatrix} = 0.098\,. \tag{16}$$

Therefore according to (16) the service system failure depending on the breakdown of the first service point is 0.098. By the similar way the probability of this system failure depending on the breakdown of the second service point is calculated and this probability is $p\begin{pmatrix}1\to 0\\ x_2\\ 1\to 0\end{pmatrix} = 0.098$.

There are three boundary states for the system failure depending the fault of infrastructure (the third component): $p_{(01)}\begin{pmatrix}1\to 0\\ x_3\\ 1\to 0\end{pmatrix} = 0.144$, $p_{(10)}\begin{pmatrix}1\to 0\\ x_3\\ 1\to 0\end{pmatrix} = 0.084$ and $p_{(11)}\begin{pmatrix}2\to 0\\ x_1\\ 1\to 0\end{pmatrix} = 0.336$. Therefore the probability the service system fault caused by the failure of the infrastructure is 0.564.

Other probabilities of this system failure or deterioration of the performance level are calculated according to (10) – (14) similar.

5 Conclusion

The mathematical background for application of mathematical methods of MVL for reliability analysis of MSS is considered in this paper. The correlation of the structure function and MVL function are shown and proved by means of the conception of incompletely specified MVL function. This background allows using Direct Partial Logic Derivatives for analysis of MSS structure function.

Mathematical approach of Direct Partial Logic Derivatives in MVL is used for investigation of dynamic properties of MVL function. The analysis of boundary values of MVL function is possible based on these derivatives too. In this paper the investigation of boundary values of the structure function of MSS and definition of MSS exact boundary states based on these valued are considered. Conception of exact boundary states is defined as boundary states for fixed MSS performance level change depending on change of the appointed change components state. This conception is extended for exact boundary states depending on changes of some components states.

New measures for estimation of MSS boundary states estimation are introduced and considered in the paper. The analysis of MSS based on the exact boundary states has not limits for the numbers of components (n) and states for every component (m_i), and system performance levels (M) according to the theoretical background. But in real-world applications these numbers have important influence to the structure function dimension (number of structure function elements) that is calculated as:

$$N_{\text{structure function dimension}} = m_1 \times m_1 \times \ldots \times m_n$$

As a rule the number of system performance levels (M) and number of component states (m_i) are defined between two and seven. The increase of the structure function

dimension depending on the number of components (n) is illustrate in Fig.5. According to the investigation in papers [15 – 17] the Direct Partial Logical Derivatives is applicable for systems which have dimension less than ten millions elements. So the proposed method can be used for the MSS at least ten components.

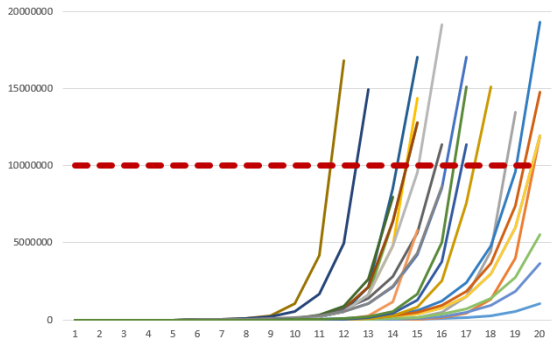


Fig. 5. Calculation of the Direct Partial Logic Derivative $\partial\phi(1\rightarrow0)/\partial x_1(1\rightarrow0)\,\partial x_3(2\rightarrow1)$.

Acknowledgments. This work was partially supported by grant of Scientific Grant Agency of the Ministry of Education of Slovak Republic (Vega 1/0498/14) and the project "Innovation and internationalization of education - instruments for increasing quality of the University of Žilina in the European educational space".

The publication was created as part of the project: "Innovation and internationalization of education - instruments for increasing quality of the University of Žilina in the European educational space"
Modern education for the knowledge society / project is co-financed from EU funds
Európska únia

References

1. Zio, E.: Reliability engineering: Old Problems and New Challenges. Reliability Engineering and System Safety. 94, 125–141 (2009)
2. Lisnianski, A., Levitin. G.: Multi-State System Reliability. Assessment, Optimisation and Applications. World scientific (2003)
3. Murchland, J.D.: Fundamental Concepts and relations for Reliability Analysis of Multistate System. In: Reliability and Fault Tree Analysis, Theoretical and Applied Aspects of System Reliability. SIAM, pp.581-618. (1975)
4. Barlow, R.E., Wu, A.S.: Coherent System with Multi-State component. Mathematics of Operations Research. 3, 275–281 (1978).
5. Hudson, J. C., Kapur, K. C.: Modules in Coherent Multistate Systems. IEEE Trans. Reliability. 32, 183–185 (1983)
6. Xie, M., Dai, Y.-S., Poh, K.-L.: Multi-State System reliability. In: Computing System Reliability. Models and Analysis, New York, Kluwer Academic Publishers, pp.207–237 (2004)
7. Levitin, G., Lisniansi, A.: Optimization of Imperfect Preventive Maintenance for Multi-State System. Reliability Engineering and System Safety. 67, 193–203 (2000)

8. Zio, E., Marella, M., Podofillini, L.: A Monte Carlo Simulation Approach to the Availability Assessment of Multi-State Systems with Operational Dependencies. Reliability Engineering and System Safety. 92, 871–882 (2007)
9. Yu, k., Koren, I., Guo, Y.: Generalized Multistate Monotone Coherent Systems. IEEE Trans Reliability. 43, 242–250 (1994)
10. Caldarola, L.: Coherent System with Multi-State Components. Nuclear Engineering and Design. 58, 127–139 (1980)
11. Veeraraghavan, M., Trivedi, K.S.: A Combinatorial Algorithm for Performance and Reliability Analysis Using Multistate Models. IEEE Transactions on Computers. 43, 29–233 (1994)
12. Hudson, J. C., Kapur, K. C.: Modules in Coherent Multistate Systems. IEEE Trans. Reliability. 32, 183–185 (1983)
13. Reinske, K., Ushakov, I.: Application of Graph Theory for Reliability Analysis. Moscow, Radio i Sviaz (1988) (in Russian)
14. Zaitseva, E.: Reliability analysis of Multi-State System. Dynamical Systems and Geometric Theories. 1, 213–222 (2003)
15. Zaitseva, E.: Importance Analysis of a Multi-State System Based on Multiple-Valued Logic Methods. In: Lisnianski, A., Frenkel, I.(eds) Recent Advances in System Reliability: Signatures, Multi-state Systems and Statistical Inference, pp. 113-134. Springer: London (2012)
16. Zaitseva, E., Levashenko, V.: Multiple-Valued Logic Mathematical Approaches for Multi-State System Reliability Analysis. Journal of Applied Logic. 11, 350–362 (2013)
17. Kvassay, M., Zaitseva, E., Levashenko, V., Minimal Cut Sets and Direct Partial Logic Derivatives in Reliability Analysis, In: Safety and Reliability: Methodology and Applications – Proceedings of the European Safety and Reliability Conference, pp. 241–248. CRC Press (2014)
18. Nahas, N., Nourelfath, M.: Ant System for Reliability Optimization of a Series System with Multiple-Choice and Budget Constraints. Reliability Engineering and System Safety. 87, 1–12 (2005)
19. Miller, M.D., Thornton, M.A. Multiple Valued Logic: Concepts and Representations. Synthesis Lectures on Digital Circuits and Systems. Morgan & Claypool Publishers (2008).
20. Vachtsevanos, G., Lewis, F. L., Roemer, M., Hess, A., Wu, B.: Intelligent Fault Diagnosis and Prognosis for Engineering Systems. Hoboken. John Wiley and Sons, NJ (2006)
21. Ramirez-Marquez, J.E., Coit, D.W., Tortorella,M.: A Generalized Multistate Based Path Vector Approach for Multistate Two-Terminal Reliability. IIE Transactions. 38, 477–488 (2006)
22. Yeh, W.C.: A Fast Algorithm for Searching All Multi-State Minimal Cuts. IEEE Trans Reliability. 57, 581–588 (2008)
23. Boedigheimer, R.A., Kapur, K.C.: Customer-Driven Reliability Models for Multistate Coherent Systems. IEEE Trans Reliability. 43, 46–50 (1994)
24. Zaitseva, E., Kovalik, S., Levashenko, V., Matiaško, K.: Algorithm for Dynamic Analysis of Multi-State System by Structure Function. In: IEEE International Conference on Computer as a tool, pp.1224–1227. IEEE Press (2005)
25. Zaitseva, E.: Dynamic Reliability Indices for Multi-State System. In: the 33th IEEE Int. Symp. on Multiple-Valued Logic, pp. 287-292. IEEE Press (2003).
26. Kapur, K.C., Zaitseva, E., Kovalik, S. , Matiasko, K.: Customer-Driven Reliability Models and Logical Differential Calculus for Reliability Analysis of Multi State System. Journal of KONBiN. The 4th Int. Conf. on Safety and Reliability. 1, 39–47 (2006)
27. Hung, J., Zuo, M.J.: Multi-State k-out-of-n System Model and its Applications. In: Ann. Reliability & Maintainability Symp., pp. 264–268. IEEE Pres, New York (2000)

Automation of Building the Safety Models of Complex Technical Systems for Critical Application

Bohdan Volochiy[1], Bohdan Mandziy[1], Leonid Ozirkovskyy[1]

[1] Department of Theoretical Radio Engineering and Radio Measurement, Lviv Polytechnic National University, 12 Bandera str., 79013 Lviv, Ukraine
bvolochiy@ukr.net, bmandziy@lp.edu.ua, l.ozirkovsky@gmail.com

Abstract. In this paper the improvement of method of automated building of state space models of complex technical systems for critical application was proposed. On the basis of the developed model with the split state of critical failure the reliability and safety indexes of studied system can be obtained. Developed approach allows to estimate of reliability and safety indexes, that makes the impact of maintenance strategies on safety and reliability, impact of the fault tolerance on safety to be considered. This will increase the accuracy (certainty) of efficiency indexes estimation of complex technical systems for critical application.

Keywords. Reliability, Reliability Engineering, Safety, Modeling, Complex System.

Key Terms. Reliability, MathematicalModel, MathematicalModeling

1 Introduction

Modern technical systems belong to the class of complex systems, which have the following properties [1,2]:

- presence of large number of elements which interact according to the given fuctional algorithm, that causes the great dimention of reliability mathematical model (from tens to hundreds of thousands of differential equations);
- elements of the system can be characterized by several types of failures (such as breakage and short circuit);
- in the case of multifunctional systems the situations when not all functions are fully performed or are performed simultaneously or are performed with the deterioration of relevant characteristics can happen. Therefore the definition of "system failure" is complicated;
- the failure of complex system for critical application can affect the human casualties or material damage, so these systems must be fault-tolerant (the ability to function normally in conditions of failures of individual elements) and safety (resistance to accidents). These properties are achieved by introduction of various kinds of redundancy (structural, algorithmic, time,

etc.), that leads to complexity of structure and internal behavior algorithm of the system as a result of the introduction of control functions, fault isolation and function recovery.

Thus, the designers of complex technical systems for critical application (CTSCA) must provide the high level of reliability and safety of the project, and thus they meet a number of contradictions, namely: contradiction between the complexity of the system and its reliability (more complex system has lower reliability), the contradiction between reliability and safety (to increase the level of safety it is necessary to induce additional subsystem of security, locking, emergency stop, etc., that reduces reliability). Increasing of reliability through the use of fault-tolerant configuration does not increase the level of safety. At the same time the applying of appropriate maintenance strategy increases both reliability and safety. At the design stage, these contradictions are solved by multivariate mathematical modeling of CTSCA, comparative analysis of alternatives and selection of the best one. Note that the system reliability analysis involves the study of the process of transition from state to state in the state space as a result of failure or restoration of certain elements of the system. In the general approach for forming reliability models these models are formalized and describe the interaction of elements of the system while its performing from the reliability position. These models reflect the degree of each element influence on the reliability in the whole. The study of safety includes, in addition, the analysis of the transition of system failures due to accident and determines the characteristics of this process.

Due to complexity of modern technical systems the multivariate analysis without automation of model building and estimation of reliability and safety indexes on its basis are not available in many cases. So often, especially for safety estimation, it is replaced by building one variant of the model followed by the combination of obtained results with expert evaluation of safety and recommendations to bring them up to acceptable values.

Nowadays reliability behavior modeling of CTSCA and its safety modeling are carried out independently of each other, using different types of models, which in the case of reliability take into account some properties of the system, but in the case of safety – completely different, although in reality these properties are interrelated and can’t be separated.

This approach is explained by the reliability models complexity as well as safety models and respectively by huge time costs for their building and by significant computational costs for their analysis with taking into account only the important nuances of CTSCA behavior. The dimension of reliability models of modern systems can reach hundreds and thousands equations. The safety model is, unlike the reliability model, complex logical function that contains hundreds and thousands arguments. Experience shows that the "manual" building of reliability models of fault-tolerant systems even with small number of elements (10) without software usage requires time-consuming procedure of dozen hours. If you change the parameters of the state graph you need to rebuild the new one and the probability of making errors in the model is very high when the chances of detection them is very low, also the time of restructuring the state graph is comparable with the time of construction its first version. Manual building of safety models as fault tree and the risk indexes estimation on its basis

(minimal cut set) is comparable to the complexity of the building the reliability models as graph of states and transitions.

From the above it arises the urgent task of further improvement and development of automated methods for modeling reliability behavior of CTSCA which are focused on reliability and safety indexes estimation.

2 The Current State of Modeling the Reliability and Safety of Complex Technical Systems Critical Application and Directions for Its Improvement

For reliability estimation of CTSCA nowadays there are enough formal and in some cases software implemented approaches, but for safety estimation there are only partially formalized methodologies which involve manual building of logical and probabilistic models in GUI. These models provide the automated determination of selected safety indexes - risk indexes (minimal cut sets).

Well-known software suites such as RAM Commander (ALD, Israel) [3], PTC Windchill QualitySolutions (PTC, USA) [4], ReliaSoft Synthesis Master Suite (ReliaSoft USA) [5], Item Toolkit (Item Software, USA, UK) [6], Reliability Workbench (Isograph, US, UK) [7] allow building reliability models as reliability block diagrams (RBD) with the automated estimation of reliability. Models as graph of states and transitions are built manually with further automation of reliability analysis.

For safety estimation these software suites have graphical tools for forming fault trees in manual mode with the automated determination of minimum cut sets and special tools to carry out FMEA / FMECA analysis. The main advantage of these software suites is that they contain integrated frameworks of elements models (electronic, electromechanical, mechanical, etc.) in accordance with international standards: MIL-HDBK-217, Telcordia SR-332, IEC TR 62380, 217Plus, FIDES, which are required for reliability and safety analysis.

In monograph [1] the general principles of automation of building reliability models as matrix of states and transitions and matrix with subsequent transition to the graph of states and transitions are given as guidelines and recommendations. Also, this approach does not have tools to analyze safety. In monograph [8] the fundamental principles of logical and probabilistic models as fault trees for the reliability and safety estimation are provided. Actually, this approach is widely used to analyze safety indexes, namely, risk by the minimal cut sets determination. However, this approach isn't formalized and in the case of CTSCA it requires significant time costs for building the fault tree and computational costs for the analysis of safety indexes. In addition, any changes in the structure of the system require the construction of its new model. Therefore, for multivariate analysis at the design stage this approach is rarely used, it is usually provided for certification, when the structure of CTSCA is established.

Currently the most powerful method for building reliability models of CTSCA is the state space method. It allows us to adequately reflect the functional and reliability behavior of CTSCA. Generated by this method model is represented by yhe system of linear differential equations of Chapman-Kolmogorov which adequately describes all

the features of system behavior that allows us to obtain standardized and non-standardized reliability indexes, which are required by developer at design stage. However, for the analysis of safety and risk, in particular, this mathematical tool is not used in practice, although there are attempts to use it for building dynamic fault trees [10]. Practical use of state space method [11] is limited at the design stage, due to cumbersome models, the phase space of which is equal to 103 .. 104 equations, and for multivariate analysis, in most cases, it is replaced by simplified evaluation using standard models.

In work [2] the method of automated generation of state space for behavior analysis of CTSCA basing on formalized description of the designed object in the form of structural-automatic model is described. It allows us to automate the process of reliability models building and to significantly reduce the time costs of multivariate analysis.

Structural-automatic model (SAM) consists of three sets of data [2]. The first set is state vector (SV), which describes all the formalized list of states using variables - SV components. The SV components are variables which describe the state of the system elements. State vector may contain additional components which are used to track the status of additional features, such as counter of current number of repairs of each item; counter of all repairs; counter of total number of failed items and so on.

The second set is constants - set of formal parameters which characterize the structure of the system and its properties, namely the number of parts on the system configuration, the number of reserve elements, their failure rate and intensity of renewals, limited number of updates and more.

The third set is tree of modification rules of state vector components (TMRSVC), which is given in tabular form and reflects the consequences which come after the failure or recovery of certain elements under certain conditions. The components of TMRSVC are the events, which can occur with elements (failure or recovery of element, reserve connection etc.), the set of logical conditions that defines combinations of values of state vector components, which take place for this event, and the modification rules of states vector components (MRSV). Each condition corresponds to the formula for calculating the intensity transition (FCIT). The event result is the change of SV component and transition from one state of system to another in accordance with the rules of transition. If certain elements inherent in more than one type of failures (such as breakage and short circuit), the probability of which is known, in such cases, use the set of formulas for calculate the probability of alternative transitions (FCPAT), for each of which the certain rule from MRSV is used.

Time-costs for build the SAM by experienced developer are 1-30 hours, depending on the complexity of the system. These costs justifies itself in multivariate analysis of fault-tolerant systems, because the next correction of the model, even with significant changes in the structure of the system takes time from tens of minutes to several hours.

This approach is implemented in ASNA software[2, 11]. Input data about the researched object for software module ASNA should be submitted in the form of SAM, which is formalized description of the structure and reliability behavior of system (the rules of transition from one state to another during the failure and recovery of elements). Basing on SAM software module ASNA generates the list of all possible states of the

system, the table of transitions from one state to another, which is transformed into the matrix of intensities of transitions when entering numerical values of intensities of failures and recovery of the system. Therefore, basing on the matrix of intensities ASNA module automatically forms the system of differential Chapman-Kolmogorov equations and solves it by Runge-Kutta-Merson method. As a result the user gets the time dependences of probabilities of system being in each of the possible states. Basing on this information, the user can define standardized reliability indexes of system (availability function, probability of failure, failure flow parameter, MTTF, etc.), and arbitrary parameters that may be needed for the "thin" study of the system (probability of downtime, probability of having at least N employable elements when using a certain number of renewals, etc.).

This approach focuses on estimation reliability indexes for reliability design and efficiency indexes for functional design. To use this approach to the safety indexes estimation the improvement both the graph states and transitions (to display emergency situations) and description of the state vector and principles of SAM building is needed. In particular, in work [12] it is proposed to combine the approach outlined in the monograph [1] with reliability block diagram GUI, allowing us to integrate into SAM designed method of RBD visualization and determine system operation conditions. Developed interface allows entering data not only for method of RBD visualization, but also for reliability model of the whole system. However, this approach has several limitations considering maintenance strategies and tools for monitoring and diagnostics. In addition, this approach focuses exclusively on building reliability models.

Thus, among the known approaches there were not found ones which allow determining the reliability and safety indexes for the same behavior model of CTSCA with taking into account all behavior features of the system while disability, accidents, downtime, etc. Hence the task of updating SAM and state space method for their adaptation to the problems of multivariate analysis and safety indexes estimation.

3 Improvement of the State Space Method and Its Formalization for Safety Models Building

The state space method combining with formalized description of the systems in the form of SAM is the powerful tool for the study of both functional and reliability indexes of CTSCA STSVP that allows us to perform multivariate analysis with minimal time-cost. Significant advantages of the state space method is that it provides the set of all states of CTSCA and determine the probability rates getting in or staying in any of them. This property is particularly relevant when the operation of the system allows the states of reduced functionality or partial disability. In addition, you can see the quantity of reliability increase when entering certain types of redundancy and their cost. These properties make it possible not only to investigate the reliability of CTSCA when carrying in redundancy or changing its behavior algorithm, but also to analyze the impact of these actions on safety, which we understand as the risk of emergency in case of failure of each element of system.

This index according to [13, 14] is called minimal cut set. Minimal cut set (MCS) – is a minimal combination of events which lead to catastrophic system failure. If when any of event is removed from the MCS the remaining events collectively cannot cause to catastrophic system failure [13].

Thus, when designing CTSCA we must have a single model that is based on the state space method and provides:

- adequate reflection of system behavior while disability;
- consideration of strategies for maintenance and repair;
- consideration of controls and diagnostics;
- possibility of obtaining reliability indexes (probability of faultless work , availability, MTTF, MTBF);
- possibility of obtaining safety indexes (MCS);
- consideration of system downtime;
- the opportunity to obtain indexes of economic efficiency;
- to carry out the multivariate analysis.

To achieve this goal it is necessary to make a number of modifications of the state space method, as described below. The behavior of CTSCA is described by graph of states and transitions. Vertices of graph are the states in which the system can be. These states are characterized by probabilities. Edges of the graph are the possible transitions from state to state and are characterized by the transition intensities.

In all known methods the catastrophic failure condition is a combination of all inoperable states, which are united in one state. This is used, on the one hand, to obtain the required reliability indexes when only operable states are used, on the other hand, inoperable states significantly increase the phase space, dimensionality of which is great.

Therefore, for safety indexes estimation (MCS), you must split the state of catastrophic failure (CF) in separate states. Thus, the set of inoperable states contains a subset of accidents (AS_1, ..., AS_i, ...), accordingly to CTSCA (Fig. 1). Each of these accidents can be represented by the corresponding fault tree.

The dimension of the system of differential equations Chapman - Kolmogorov increases proportionally to the expansion of phase space and the system of equations consists of two parts - the equations that describe the operable states (P_i (t)) and the equations that describe inoperable states (Q_j (t)).

The solution of the equation system can be implemented by analytical methods (matrix exponential, Laplace transform) and numerical methods (Runge-Kutta, Rosenbrock). As a result of solution the probability distribution of CTSCA being in all states is obtained.

The next step is filtration of obtained probability distribution for the separation of states to operable and inoperable. Filter is in this case the condition of critical failure. As a result of filtration, we obtain a set of probability of CTSCA being in operable states $\{P_i (t)\}$ and the set of probability of CTSCA being in inoperable states $\{Q_j (t)\}$, where i is the serial number for operable states and j is the serial number for inoperable states.

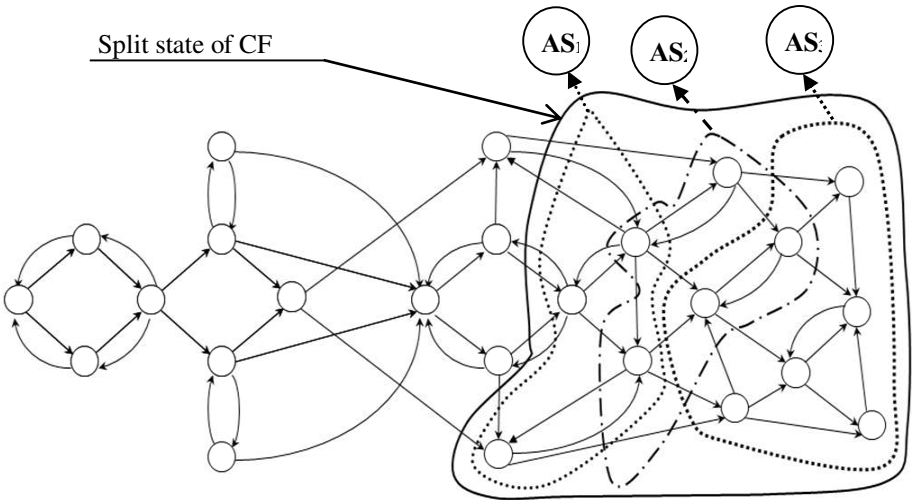


Fig. 1. Graph of state and transitions with split state of catastrophic failure

From the resulting set of operable states the necessary reliability indexes are formed and from the set of inoperable states the MCS – combination of inoperable states, when the critical failure definitely will occur – are obtained.

As the number of inoperable states is equal to $10^1..10^2$, for automated MCS obtaining, an algorithm for finding all combinations of inoperable states, which refer to critical system failure, should be developed. This means that this element is one of the most critical parts of the system. In the case of fault-tolerant systems, CTSCA is just that, the combination of several elements is possible. It is considered that as more inoperable states are included in MCS so the less vulnerable system is and so the effects of its failure will not be catastrophic for human life and health and the environment.

If vulnerable elements, which form inoperable states, which are included in MCS, are replaced by more reliable or reserved, the risk of accident is reduced in times. Thus, the MCS are necessary for designer to make reasonable redundancy in a new version of designed CTSCA. So due to the effect of redundancy input we can quantify the rate of risk reduction:

$$K_{rr}=C_m/C_n, \quad (1)$$

where

C_m – MCS before redundancy input;

C_n – MCS after redundancy input;

Generalized diagram of technique of estimation of safety and reliability indexes basing on the graph of states and transitions with the split failure state and using SAM is shown in Fig. 2. According to it, the automated algorithm for obtaining MCS was developed. The input data for the algorithm is the set of inoperable states (MCS), derived from the binary SAM.

The binary SAM is the SAM of the CTSCA, in which all elements of structure are displayed by individual SV components and can take only of two values: zero and one. The binary SAM, which, unlike to original SAM [2], makes a possibility to describe

the structure and behavior of CTS without unification of states of its structure elements. In addition, the binary SAM allows obtaining split failure state, in which states of CTSCA subsystems failures can be discerned with the given level of detail representation.

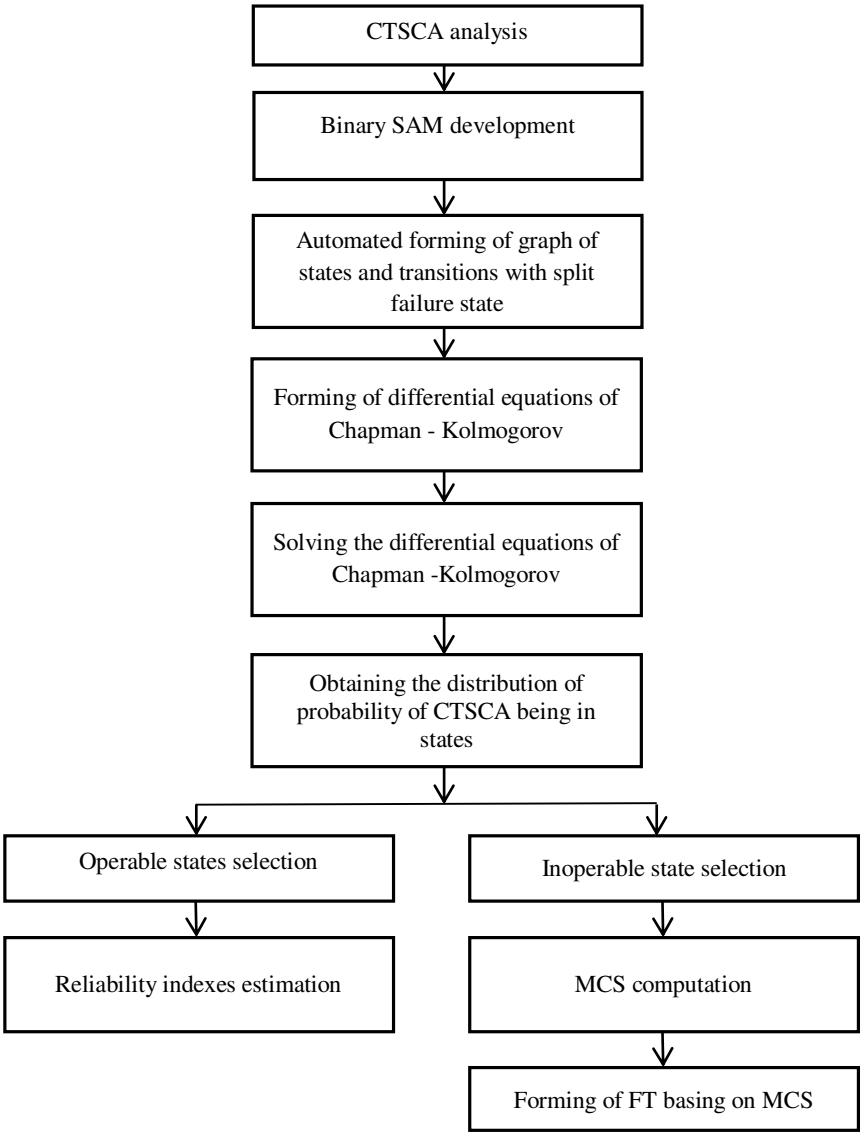


Fig. 2. Generalized diagram of technique of estimation of safety and reliability indexes basing on the graph of states and transitions with the split failure state

Procedure of filtering inoperable states from whole phase space is carried out by the analysis of the state vector component, comparing them with the critical failure

condition. If the element is operable, the value of its corresponding SV component is greater than zero. If the element failed and led to accident, the component will be equal to zero.

While the algorithm development it is taken into account that:

- at least one MCS is presented in the sysytem;
- cut set of the system is inoperable state, when system falls into catastrophic failure condition;
- MCS of the system is the state, when the system is in catastrophic failure but taking off at least one of the elements that are failed in this MCS, the catastrophic failure of the system can not occur at all.

Definition of MCS is provided in two stages: stage of MCS obtaining and stage of estimation their probability values.

Stage I. For MCS finding the following procedures are used: MCS sorting; MCS determination.

At this step it is necessary to sort obtained array of inoperable states of the system on the feature of the smallest number of events that led to the accident of the system. Further, basing on sorted array of inoperable states the MCS are defined. As a result of the proposed procedures the array of MCS is presents as a matrix.

Stage II. Determination of MCS probability is performed by the following procedures: determination of MCS from all cut sets; sum of MCS probabilities; forming of array of MCS and their probability values.

According to this stage we must create a matrix that consists of four columns – the first column is a serial number of MCS – N; the second column is SV component and its value; in the third column the numbers of states, which are attended by the corresponding MCS, are recorded. So in the fourth column there are recorded obtained probabilities of MCS as a result of this procedure. Also at this stage procedure of comparison of the system states is used.

The procedure for obtaining probability values of MS is the sum of probability values of being in respective states, whose numbers were found in the previous procedure, i.e., in the states that are recorded in the third column corresponding to the MCS matrix. As a result, the fourth column is filled with appropriate MCS probabilities value.

An example of the usage of developed method of MCS definition. Fault-tolerant system consists of five modules A, B, C, D, E. Modules A, B, D are the main operable configuration that provides performance of system functions and modules C and E are reserve modules. Modules A and B are reserved by module C. The entire system is reserved by module E. All modules have the same failure intensity $\lambda = 0{,}001$, and the observation period is $T = 100$ h.

The RBD of the fault-tolerant system is shown in Fig. 3:

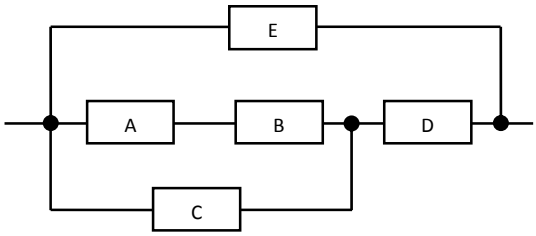


Fig. 3. The RBD of the fault-tolerant system

On the basis of developed binary SAM of the fault-tolerant system, which consists of set of formal parameters (Fig. 4), SV components and failure condition (Fig. 5), the tree of modification rules of state vector (Fig. 6), which is the input to the software module ASNA, the graph states and transitions was obtained in the automatic mode (Fig. 7).

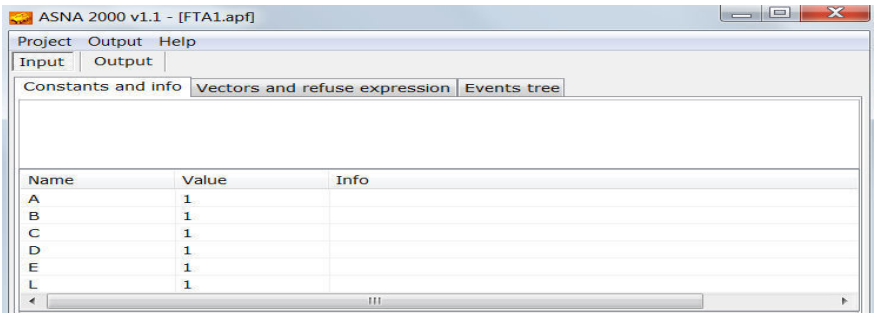


Fig.4. The set of formal parameters

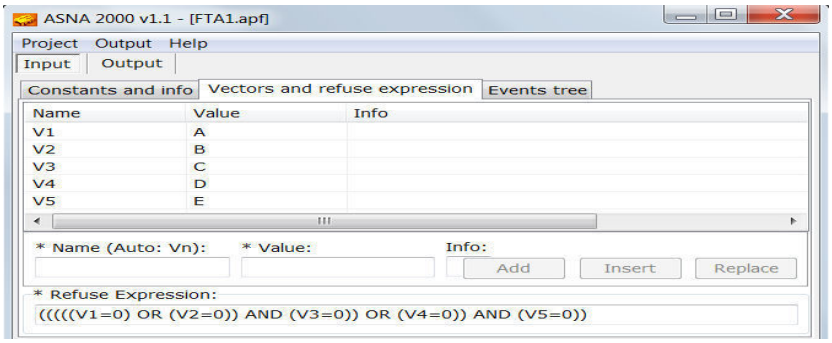


Fig .5. State vector components and failure condition

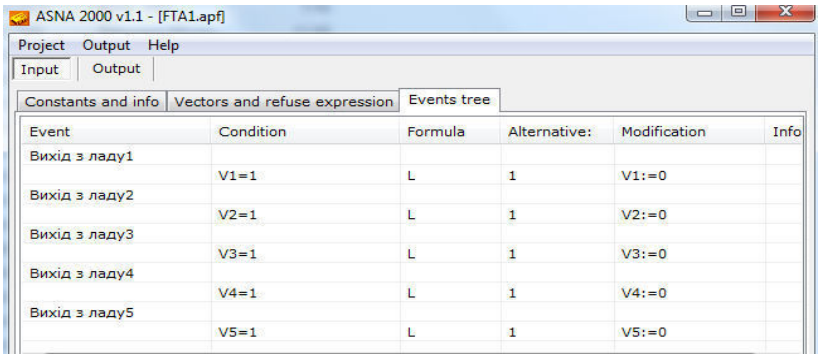


Fig.6. The tree of modification rules of state vector

Basing on the obtained graph of states and transitions the software module ASNA formed mathematical model of the system as a system of Chapmen - Kolmogorov differential equations. After its solving the probability of being in every possible state was obtained. Probability of system being in operable state is 0.9894, and the probability of failure is equal to:

$$Q_f = 1 - 0{,}9894 = 0{,}01061$$

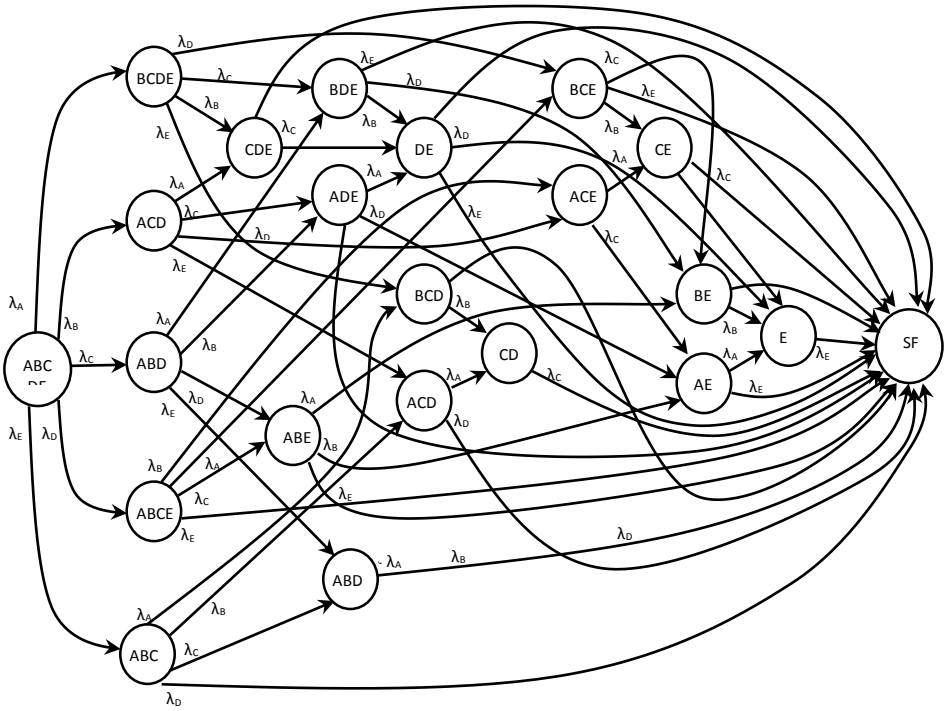


Fig. 7. Graph of state and transitions

On the basis of the graph of states and transitions according to developed algorithm, it was determined that after simultaneous failure of modules E and D the system fails

in general. Next, other two combinations which also lead to failure of the whole system are ACE and BCE. Thus, these three combinations make the MCS. The next stage was the determining of the values of the probability of each of these combinations. Substituting logical expression of MCS DE: ((V4 = 0) AND (V5 = 0)) instead of failure condition the MCS value of probability simultaneous failure of combination of modules E and D was obtained, which is Q_{DE} = 0,009. Similarly, substituting logical expression of MCS ACE and BCE instead of failure condition we get: Q_{ACE} = 0,00084; and Q_{BCE} = 0.00084. The calculated MCS is shown in Table. 1.

Table 1. Minimal cut sets of the system

Failed modules	The number of failed modules	MCS values	Percentage values, %
E, D	2	0,009	84,6
A, C, E	3	0,00084	7,92
B, C, E	3	0,00084	7,92

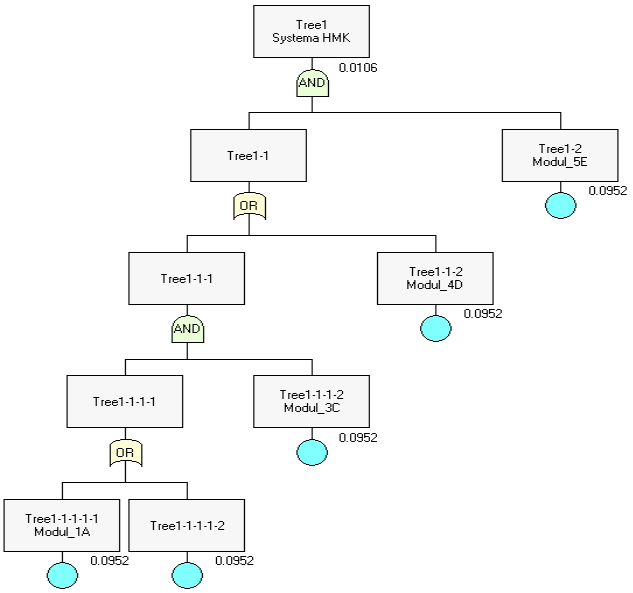


Fig.8. Fault tree

Validation of the developed method. To validate the developed method it was implemented the a fault tree building for the system (Fig. 8) according to the approach [8] and the values of the probability of failures for each MCS were calculated. It was considered that the results obtained by fault tree are accurate and they were compared with results which are shown in Table 1.

The validation was performed using specialized software suite RAM Commander by ALD Service. For RBD the fault tree was set up (Fig. 8) and MCS were obtained by tools of RAM Commander and are shown in Fig. 9.

The comparison shows that the calculated values of MCS, which were obtained from fault tree using software suite RAM Commander coincide with the values obtained from the graph of states and transitions with the split failure state using binary SAM.

The developed approach (Fig. 2) allows us to get the MCS in automatic mode without fault tree construction.

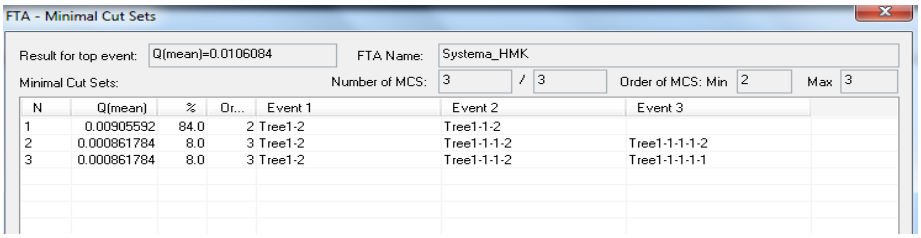


Fig. 9. Minimal cut sets basing on fault tree

4 Expanding the Functionality of the Program ASNA for the Safety Analysis of CTSCA

For building complex models, which are focused on determination of the reliability and safety indexes it is most advisable to take as a basis the graph of states and transitions with split state of catastrophic failure and method for its automated construction using binary SAM. However, the biggest problem for the designer, in this case, is the construction of the binary SAM because its formation requires from the developer not only the deep knowledge of the nuances of functionality of designed CTSCA but also thorough knowledge about techniques of construction the formalized graph of states and transitions that is the whole direction in complex systems designing.

Therefore, the next urgent task is to automate the construction of binary SAM-based graphical representation of the system as a RBD. This will speed up the development of SAM, reduce the time cost in degree and obtain both reliability and safety indexes. Principles of this automation were laid in works [12, 15]. At the same time, we note that this approach narrows the class of the analyzed systems because it does not allow us to analyze complex technical systems that are described by queuing systems, flowcharts, etc. behavior algorithm.

According to approach [12] the visualization software for RBD of technical system, which makes it possible the automatic construction of graphic images of flow diagram of technical systems and the formation of conditions of their functioning and failure, was developed. Using the developed software the information about the system is transmitted as input to the ASNA software for further calculations of reliability indexes accordingly to the number of elements in the node, the number of renewals and maintenance crews, time range, intensity of failures and recoveries for each of elements of analyzed system.

In order to extend the functionality of the ASNA software for safety analysis of CTSCA it is needed to combine binary SAM methodology with the approach [12]. It is necessary to modify the SAM as follows:

- Every element input in the RBD is accompanied by the creation the next set of SV components, the number of elements corresponds to the number of components:

$$Item_1, Item_2, \ldots, Item_i, \ldots \rightarrow V_{11}, V_{21}, \ldots, V_{i1}, \ldots$$

The initial value of each component is equal to one: $V_{i1}=1$;

- Type of connection of RBD elements (serial, parallel, combined) is given by the inoperable condition
- If the limited number of renewals of system is planned, for each item is added another SV component – counter of repairs:

  → $V_{12}, V_{22}, \ldots, V_{i2}, \ldots$

  The initial value of each component is equal to zero: $V_{i2}=0$
- If the number of renewals is unlimited, the additional component isn't added;
- Each RBD element is assigned to line of binary SAM as follows:

Event	Condition	FCIT	FCPAT	MRSV
Failure of module i	$V_{i1}=1$	λ_i	1	$V_{i1}=0$

- If the system is renewable, in addition to each RBD element, another line is assigned to binary SAM as follows:

Event	Condition	FCIT	FCPAT	MRSV
Repair of module i	($V_{i1}=0$) AND ($V_{i2}<RC_i$)	μ_i	1	$V_{i1}=1$ $V_{i2}= V_{i2}+1$

- Parametres of each element (failure rate - λ_i, the intensity of repair - μ_i, the number of repairs - RC_i etc.) is transmitted to set of formal parametres;
- Limited values of each RBD element repair, the number of repair crews, repair priority are transmitted to set of formal parametres;
- Inoperable conditions are transmitted to SAM and serves to filter the operable-bodied and inoperable states.

Thus all components of SAM can be automatically formed. Generated data can be represented as a file that is sent to ASNA software module as input data. ASNA software module enables automated obtaining of the graph of states and transitions with split failure state. Basing on the graph of states and transitions ASNA software makes it possible to assess reliability. CutSetDefiner software, basing on the graph of states and transitions, can generate MCS and basing on MCS through software [16] we can automatically get the fault tree.

5 Conclusions

1. Split of critical failure state in graph of states and transitions, in contrast to the known approaches, allows estimation of reliability and safety indexes, that makes the impact of maintenance strategies on safety and reliability, impact of the fault tolerance on safety to be considered. This will increase the accuracy (certainty) of efficiency indexes estimation of complex technical systems for critical application.
2. Minimal cut sets obtaining on the basis of the graph of states and transitions allows taking into account the interrelations of accidents directly from the analysis of system states for identification weaknesses. It gives only reasonable means for providing fault tolerance that reasonably reduces the cost of improving the system.
3. Using binary structural-automatic model allows automated obtaining of split critical failure state and reducing time costs for building the graph of states and transitions.

4. Risk reduction factor was introduced for quantitatively assess of the efficiency of improving safety by improving reliability by introducing redundancy in critical elements of complex technical systems for critical application.
5. Fault tree building from the graph of states and transitions basing on minimal cut sets takes into account the behavior of complex system that is not available when using static and dynamic fault trees
6. The combination of binary structural-automatic model and method of automated constructing of graph of states and transitions basing on reliability block diagram makes it possible to automate the procedure of building structural-automatic model of fault-tolerant renewable complex technical systems for critical application and reduce time costs by more than degree.

References

1. Polovko A.M., Gurov S.V.: Basics of reliability theory. BHV Peterburg Publ., Saint Petersburg (2006) (in Russian)
2. Yu. Bobalo, B. Volochiy, O. Lozynskyy, B. Mandziy, L. Ozirkovskyy, D. Fedasyuk, S. Shcherbovskykh , V. Yakovyna: Mathematical Models and Methods of Analysis of Radioelectronic, Electromechanic and Software Systems. Lviv Polytechnic National University Publ., Lviv (2013) (in Ukrainian)
3. RAMS (Reliability, Availability, Maintainability and Safety) Software, http://aldservice.com/en/reliability-products/rams-software.html
4. PTC Windchill, http://ru.ptc.com/product/windchill/quality
5. ReliaSoft Synthesis Master Suite , http://www.reliasoft.com/products.htm
6. Reliability Engineering Software. Products, http://www.itemsoft.com/products.html
7. Reliability Workbench, http://www.isograph.com/software/reliability-workbench/
8. Henley, Ernest J., Hiromitsu Kumamoto: Probabilistic Risk Assessment: Reliability Engineering, Design and Analysis. Wiley-IEEE Press, 2 edition, (2000)
9. Ajit Kumar Verma, Srividya Ajit, Durga Rao Karanki, Ajit Kumar Verma, Srividya Ajit, Durga Rao Karanki: Reliability and Safety Engineering. Springer Science & Business Media (2010)
10. Alessandro Birolini Reliability Engineering: Theory and Practice, Sixth Edition. Springer (2010)
11. Bohdan Volochiy, Bohdan Mandziy, Leonid Ozirkovskyi: Extending the features of software for reliability analysis of fault-tolerant systems. Computational Problems of Electrical Engineering, 2, 2, 113-121 (2012)
12. Mandziy Bogdan, Seniv Maksym, Mosondz Natalia, Sambir Andriy: Programming Visualization System of Block Diagram Reliability for Program Complex ASNA-4. In: Proc. of 13-th International Conference “The Experience Of Designing And Application Of Cad Systems In Microelectronics CADSM-2015”, Lviv-Slavsko (2015) (in Ukrainian)
13. Guangbin Yang: Life Cycle Reliability Engineering Hoboken. Wiley, N.J. (2007)
14. T. Zentis, R. Schmitt: Technical Risk Management for an Ensured and Efficient Product Development on the Example of Medical Equipment. In: Proceedings of the 23rd CIRP Design Conference ”Smart Product Engineering”, March 11th - 13th, pp. 387-398. Bochum (2013)

15. Mandziy B. A., Ozirkovskyi L.D.: Automation Of Building Reliability Models Of Redundant Restorable Complex Technical Systems. Eastern-European Journal of Enterprise Technology, № 4 (62), 2, 44-49 (2013) (in Ukrainian)
16. Volochiy B.Yu., Ozirkovskyi L.D., Mashchak A.V., Shkiliuk O.P.: Fault Tree Build Automation for Safety Estimation of Complex Technical System. In: Proc. of IV International conference “Physical and Technological Problems of Wireless Devices, Telecommunications, Nano-and Microelectronics PREDT-2014”, pp. 102-103 (2014) (in Ukrainian)

Scenario-Based Markovian Modeling of Web-System Availability Considering Attacks on Vulnerabilities

Vyacheslav Kharchenko[1], Yurij Ponochovny[2], Artem Boyarchuk[1] and Anatoliy Gorbenko[1]

[1] National Aerospace University KhAI, Kharkiv, Ukraine
V.Kharchenko@khai.edu
[2] Poltava National Technical University named after Yurij Kondratyuk, Poltava, Ukraine
pnch1@rambler.ru

Abstract. In the paper we simulate web-system availability taking into account security aspects and different maintenance scenarios. As a case study we have developed two Markov's models. These models simulate availability of a multi-tier web-system considering attacks on DNS vulnerabilities in additional to system failures due to hardware/software (HW/SW) faults. Proposed Markov's model use attacks rate and criticality as initial simulation parameters. In the paper we demonstrate how to estimate these parameters using open vulnerability databases (e.g. National Vulnerability Database). We also define different vulnerability elimination (VE) scenarios and examine how they affect system availability.

Keywords: web-system availability, security, vulnerability, Markov's models, scenario of vulnerability elimination

Key terms. MathematicalModeling, MathematicalModel, SoftwareSystems

1 Introduction

Efficient implementation and operation of multitier web-systems using COTS components depend on accuracy of security assessment and quality of attacks prevention and recovery activities. Security of web-system can be estimated by analyzing web-components vulnerabilities and predicting attacks affecting system availability and other security attributes. System availability and accessibility of the provided ser-vices depend on the used maintenance strategy. This strategy can implement various vulnerability prevention and elimination scenarios [1]. Thus, assessing web-systems availability taking into account both system failures due to HW/SW faults, and hacker attacks on components vulnerabilities is important.

To estimate system availability and security researchers develop various simulation models [1, 2]. Most of them are based on attack tree analysis [3,4], Markov's [5,6] and

semi-Markov's chains [7,8] or use of Petri nets [9,10] as a mathematical apparatus. However, known models do not explicitly consider attacks on system vulnerabilities causing inaccessibility of the provided services (accessibility vulnerabilities) and do not take into account different security policies and vulnerability elimination strategies.

In the paper we analyze web-system availability considering failures caused by HW/SW faults as well as attacks on system vulnerabilities. With this purpose we propose and examine a set of Markov's availability models implementing different scenarios of vulnerability elimination. This paper continues research described in [6] using scenario-based approach.

The rest of the paper is organized as follows. In the second section we suggest a set of scenarios to assess web-system availability taking into account different vulnerability elimination procedures. In the third section we discuss a technique of estimating input parameters of Markov's models by use of information about software component vulnerabilities from the open vulnerability databases. The forth section presents a case study and the set of Markov's models and also examines simulation results.

2 The Scenario-Based Approach to Web-System Availability Modeling with Regards to System Vulnerabilities and their Elimination

Attacks on vulnerabilities of web-systems can be simulated using Markov's models [5-7]. However, for that we should take into account that parameters of the vulnerabilities (numbers and types) are changed as a result of elimination and patching procedures.

In the Fig. 1 we propose a set of common state-transitional models capturing different attack and recovery scenarios. The scenarios are differed by a number of attacked vulnerabilities: one (a-f) or several (g); with (b-g) or without (a) vulnerability elimination; with vulnerability elimination after system been successfully attacked (b-d) or during (e,f) preventive maintenance actions.

We have marked model states as following: double circles correspond to up-states, single line marked circles correspond to maintenance states, thick line marked circles correspond to down-states after attacks.

The simplest scenario is shown in Fig. 1,a. After successful attack a web-system is recovered (e.g. rebooted) without vulnerability elimination. However not all attacks can be successful and lead to web system unavailability. This is why we consider two transitions from up-state S0: the first transition with the rate $\lambda attack*Da$ leads to down (unavailable)-state Sd; the second one with the rate $\lambda attack*(1-Da)$ returns back to up-state S0 (Da is a probability of attack to be successful).

The second scenario (Fig. 1,b) illustrates vulnerability elimination during system recovery after successful attack. We assume that during recovery action it is possible to eliminate from 0 to all (nv) vulnerabilities. Hence, web-system may return from the down-state Sd to the initial state S0 without vulnerability elimination with the rate $\mu'a*(1-Dp)$, where Dp is a probability of successful recovery and vulnerability elimination, or may transit to the next up-state Su with the rate $\mu'a*Dp$.

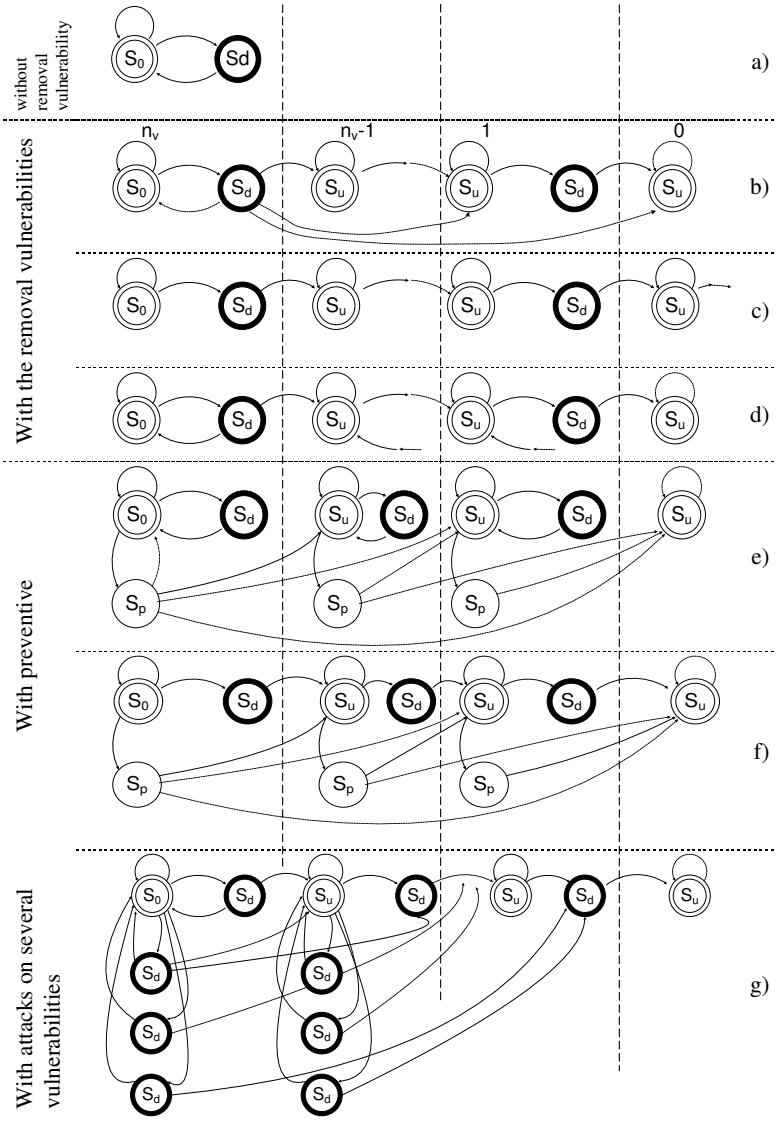


Fig. 1. Graph models of scenarios of web-system availability considering different options of vulnerability elimination

The third scenario (Fig. 1,c) describes graduate vulnerability elimination only after successful attacks on these vulnerabilities. In this scenario the total number of vulnerabilities in the system may be unlimited $nv \to \infty$.

The step by step vulnerability elimination is described by the next scenario (Fig. 1,d). In this case it is assumed that restart of web-system is possible without elimination of vulnerability which was attacked.

According with the fifth scenario (Fig. 1,e) vulnerabilities can be detected and eliminated from the system only during the periodic maintenance actions (i.e. security audits) only. After the successful attack a web-system is restarted or reboot without vulnerability elimination. Vulnerabilities can be detected and eliminated from the system only during periodic security audits. The probability of eliminating the i-th vulnerability is equal to αi, $\Sigma \alpha i = 1$.

The sixth scenario (Fig. 1,f) assumes that vulnerabilities can be detected and eliminated from the system both after successful attacks or during periodic security audits. The seventh scenario takes into account possibility of attacks on several vulnerabilities (Fig. 1,g). The scenario describes sequential chains of attacks on sever-al (four, in our example) services of a web-system. In this case an intruder continues to attack the next services. After successful attack a web-system can transit to a new up-state where vulnerabilities are eliminated from the system or can return back to the initial state by system restarting or rebooting.

Described set of scenarios is not complete. This set includes some basic scenarios. However, other scenarios can be developed considering different procedures of maintenance and vulnerability elimination or patching.

3 Estimation of Input Parameters for Markov's Web-System Availability Models

3.1 Vulnerabilities Sampling

In this section we discuss how parameters of Markov's models simulating web-system availability can be estimated using existing vulnerability databases like NVD.

The whole set of vulnerabilities stored in NVD can be downloaded as an XML file «NVD/CVE XML Feed with CVSS and CPE mappings (version 1.2)» [11,12]. Then we need to select those vulnerabilities of Web-system components (DNS-server, HTTP-server, application server, etc.) affecting system availability. It is can be done by analyzing vulnerabilities availability impact and vector of access using, for instance, common vulnerability scoring system (CVSS) [13] provided by NVD:

- Availability impact, A, which can be equal one of three fuzzy values "None" (N), "Partial" (P) and "Complete" (C);
- Vector of access, value "Network" (N).

For example, Table 1 presents a subset of vulnerabilities detected during 2013 and causing unavailability of DNS (CVSS_vector – contains – AV:N, A:C и A:P; ns1:descript – contains – DNS (an example for analysis attacks on DNS) including their publishing dates and score.

3.2 Estimation of Attack Rates

In order to parameterizes state-transition models we need to evaluate a rate of the attacks exploiting system vulnerabilities.

This rate obviously depends on different factors including number of system

vulnerabilities, their criticality, availability impact and vector of access. However, vulnerabilities define only the capability of a system to be attacked. On the other hand, unlike random system failures, vulnerabilities are exploited by various intended (hacker, computer criminals, industrial espionage, insiders, etc.) and unintended (viruses, worms, malware, etc.) threat agents.

Table 1. Subset of vulnerabilities causing DNS unavailability (01.2013 – 10.2013)

#	name	published	base_score	CVSS_vector
1	CVE-2013-0198	05.03.2013	5,0	(AV:N/AC:L/Au:N/C:N/I:N/A:P)
2	CVE-2013-2266	28.03.2013	7,8	(AV:N/AC:L/Au:N/C:N/I:N/A:C)
3	CVE-2013-2494	28.03.2013	4,9	(AV:N/AC:H/Au:S/C:N/I:N/A:C)
4	CVE-2013-1152	11.04.2013	7,8	(AV:N/AC:L/Au:N/C:N/I:N/A:C)
5	CVE-2013-2052	09.07.2013	5,1	(AV:N/AC:H/Au:N/C:P/I:P/A:P)
6	CVE-2013-2053	09.07.2013	6,8	(AV:N/AC:M/Au:N/C:P/I:P/A:P)
7	CVE-2013-2054	09.07.2013	5,1	(AV:N/AC:H/Au:N/C:P/I:P/A:P)
8	CVE-2013-4854	29.07.2013	7,8	(AV:N/AC:L/Au:N/C:N/I:N/A:C)
9	CVE-2013-4115	09.08.2013	7,8	(AV:N/AC:L/Au:N/C:N/I:N/A:C)
10	CVE-2013-5479	27.09.2013	7,8	(AV:N/AC:L/Au:N/C:N/I:N/A:C)
11	CVE-2013-5480	27.09.2013	7,8	(AV:N/AC:L/Au:N/C:N/I:N/A:C)

Motivation of intended threat agents is also depended on the system itself (its value and interest for the attacker). Last two factors are really difficult to define quantitatively. Thus, in the paper we propose to define the attack rate by the average per year frequency of vulnerability disclosure in the system components.

Criticality of attack is determined as an average value of basic CVSS estimation. We propose the following technique to estimate attack rate:

1) development of availability block diagram (ABD) of web-systems as a sequentially-parallel connection of components influencing on accessibility (similar to RBD);

2) extraction from NVD the vulnerability subsets for all components of ABD;

3) calculation of average per year frequency of vulnerability disclosure in these subsets;

4) determination of attack rate as the maximum of these frequencies of vulnerability disclosure;

5) calculation of attack criticality as an average value of basic CVSS estimation for selected set per year.

According with Table 1, average attack rate on DNS vulnerabilities causing unavailability could be estimated in 2013 as 1,26*10–3 1/h while the average criticality equals 6,75.

4 Web-System Availability Models for Different Vulnerability Elimination Scenarios

4.1 Initial Model and its Parameters

Let us examine a web-system based on three network services: DNS, DHCP and Routing. Reliability block diagram (RBD) and Markov's model (the marked Markov's chain) of the web-system are shown in Fig. 2.

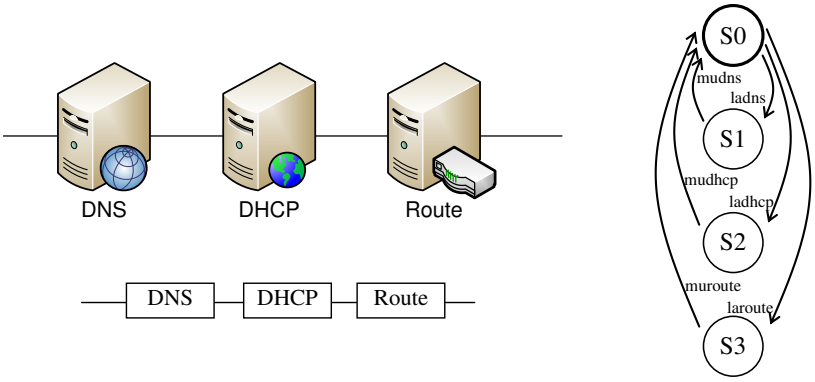


Fig. 2. Reliability block diagram and Markov's model of the web-system without considering system vulnerabilities

Table 2. Values of input parameters for availability models

#	Name	Symbol	Value	Unit
1.	DNS service software failure rate	*ladns*	3e-5	1/hr
3.	DHCP service software failure rate	*ladhcp*	1.5e-5	1/hr
4.	Route service software failure rate	*laroute*	5e-4	1/hr
5.	DNS service software recovery rate	*mudns*	0.67	1/hr
6.	DHCP service software recovery rate	*mudhcp*	1	1/hr
7.	Route service software recovery rate	*muroute*	0.33	1/hr
8.	Attack rate on availability (accessibility) of DNS service	*laatdns*	6.3e-3	1/hr
9.	Criticality of attack on availability of DNS service	*d1dns*	0.77	
10.	Restart (recovery) rate after attack on availability	*mureboot*	0.5	1/hr
11.	Restart (recovery) rate after attack on availability with VE	*murecovery*	0.22	1/hr
12.	Rate of maintenance (security audit)	*laprof*	4.5e-4	1/hr
13.	Recovery rate of service after security audit	*muprof*	0.5	1/hr
14.	Probability of successful recovery with VE	*d2p*	0.5	
15.	Probability of vulnerability elimination during security audit	p ($p=\alpha_1$)	0.7	

The RBD consists of three consequently connected components and failure of any components causes failure (unavailability) of the system. In this section we study two

availability models taking into account attacks on DNS vulnerabilities and different maintenance operations including security audits [7]. The first model (MA-1) corresponds to scenario with vulnerability elimination during security audits only (Fig. 1,e). The second one (MA-2) implements scenario with vulnerability elimination after successful attack on a system and also during security audits (Fig. 1,f).

Initial values of model parameters are presented in Table 2. The models itself have been implemented as Matlab programs.

4.2 The Model MA-1

This model describes a web-system with attacks on DNS vulnerabilities and periodic maintenance activities (security audits) including detection and elimination of vulnerabilities without complication of code (*ladns* =const).

Table 3. Probabilities of detection of j vulnerabilities

j	1	2	3	…	nv–1	nv
α_j	p	q*p	q^2*p	…	q^{nv-2}*p	$1-\sum\alpha_j$

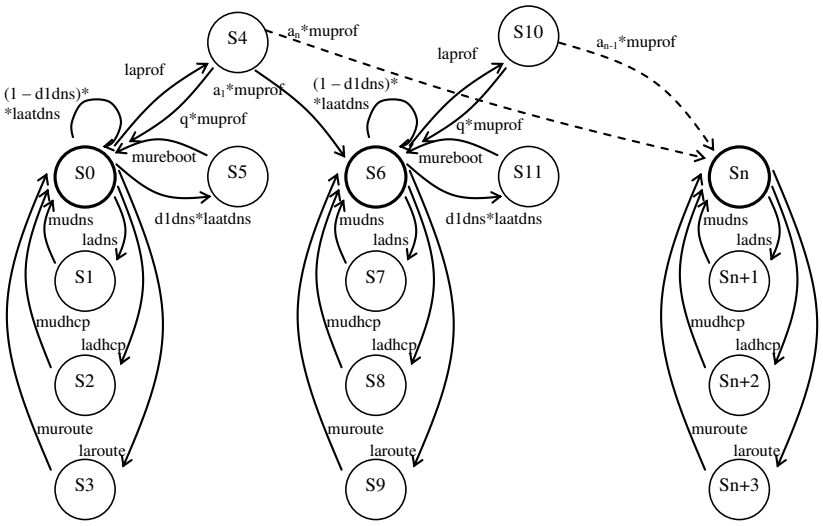


Fig. 3. Marked Markov's graph for MA-1

Marked Markov's graph is shown on Fig. 3. As during these activities it is possible to detect and eliminate more than one vulnerability [1…nv], we use a special parameter αj which defines probability of detection of j-th (j ∈ [1…nv]) vulnerabilities. Apparently, Σαj = 1, and values α1, α2,… αj,… αnv are distributes on discreet law. For calculation of geometrical distribution law αj was used with parameters: p=α1=0.7 (probability of detection of the single vulnerability) and q=1–p=0.3 (Table 3).

Initially (state S0) web-system works considering failures and recovering of DNS, DHCP и Routing services (states S1- S3). After attack on DNS (transition to state S5

with the rate d1dns*laatdns) the system fails and can be recovered by restart without vulnerability elimination with rate mureboot. Periodically maintenance activities are performed (state S4) during which 0, 1,…nv vulnerabilities can be eliminated (transitions from state S4 to states S0, S4… Sn). These transitions are weighted using parameter αj*muprof. Further process is continued in the same way (states Sn…Sn+3).

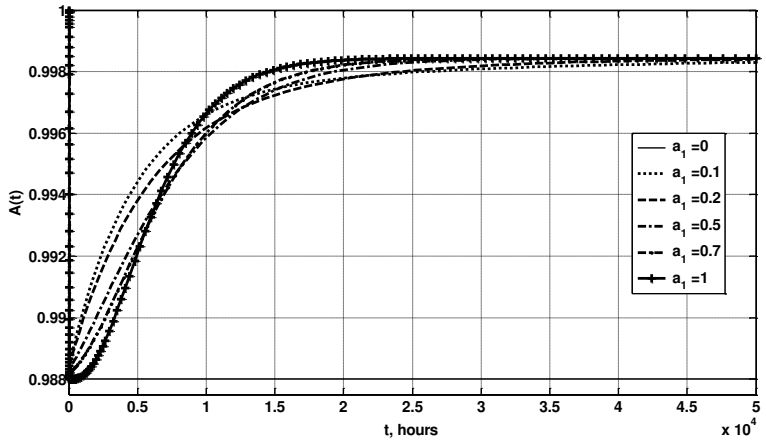


Fig. 4. Diagram of dependency of availability function for the model MA-1 on different probabilities α_1

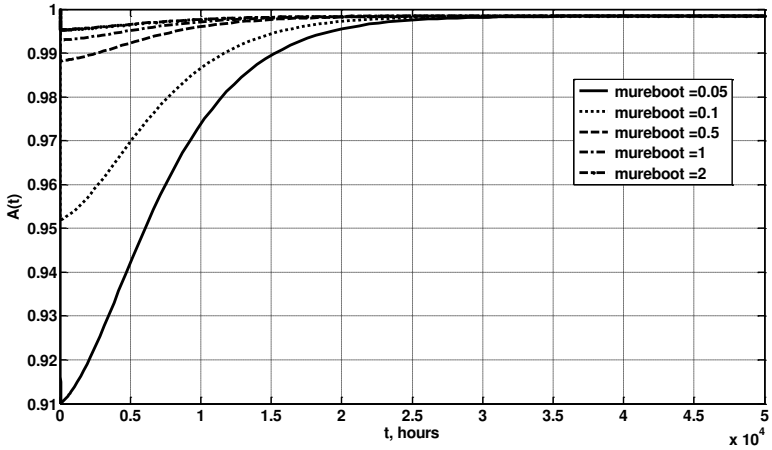


Fig. 5. Diagram of dependency of availability function for the model MA-1 on different recovery rate after attack on vulnerabilities, *mureboot*

The research results of availability function depending on parameters p=α1 and mureboot are shown on Fig. 4 and Fig. 5.

The greater value α1 causes more fast transition of the function A(t) to stationary state (Fig. 4). A value of mureboot influences on a value of availability function minimum, location of minimum on the time axis and time of transition to stationary

state (Fig. 5). If mureboot=2 (1/hour) availability function minimum equals 0,9953 for t=17 hours; if mureboot=0.05 (1/hour) availability function minimum equals 0,9103 for t=119 hours.

4.3 The model MA-2

This model describes scenarios whish in addition to MA-1 assumes detection and elimination of vulnerabilities both during security audit and right after attack (without complication of code (ladns =const). Marked Markov's graph is shown on Fig. 6.

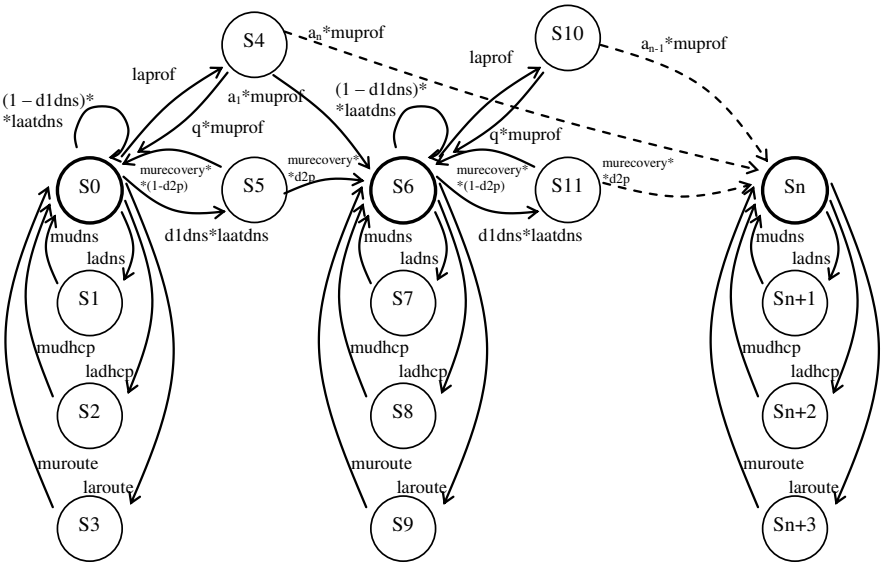


Fig. 6. Marked Markov's graph for MA-2

After attack on DNS and transition to state S5 with rate d1dns*laatdns system fails and can be recovered by restart without eliminating vulnerability with the rate (1–d2p)*murecovery or with elimination with the rate d2p*murecovery.

The results of availability function analysis depending on parameters d2p and laprof are shown on Fig. 7 and Fig. 8. The increasing of probability of vulnerability elimination d2 during maintenance activities causes more fast transition of the function A(t) to stationary state (Fig. 8). Changing the availability function depending on the rate of maintenance laprof is dual. On the one hand the rare maintenance activities are carried on the more minimum of availability function on non-stationary phase. On the other side the more often maintenance activities are carried on the faster the function transits to stationary state (Fig. 8).

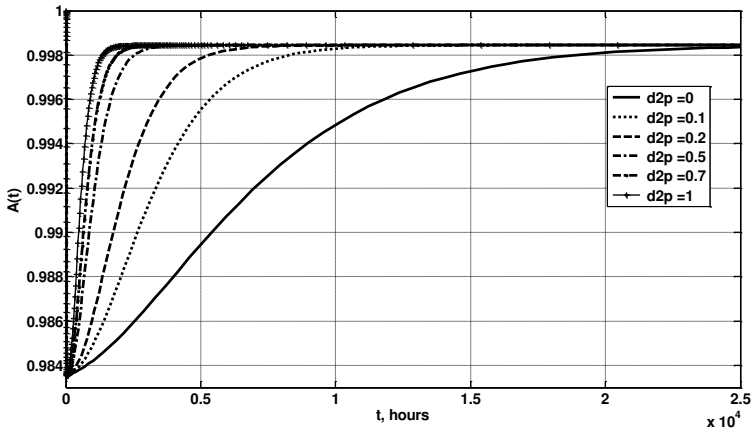


Fig. 7. Diagram of dependency of availability function for the model MA-2 on different probabilities of vulnerability elimination after attack *d2p*

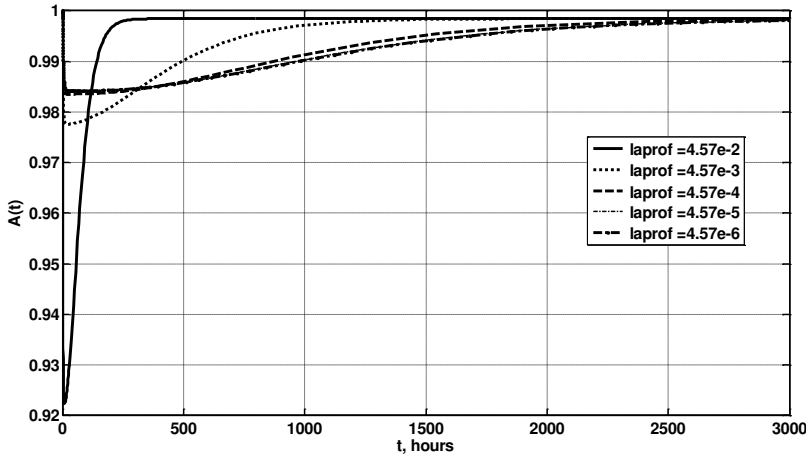


Fig. 8. Diagram of dependency of availability function for the model MA-2 on rate of maintenance *laprof*

4.4 Combining of the Models MA-1 and MA-2

The scenarios corresponding to the models MA-1 and MA-2 can be superposed to increase availability due to increasing of minimum and duration of system transition to the stationary state of availability function. To combine these two scenarios we have developed a set of Matlab programs. Filing of coefficient matrixes was done according with the same initial data (Table3). To solve systems of Kolmogorov-Chapman's differential equations the method *ode15s* for time span [0…20000] hours. The results of solving are shown on the Fig. 9.

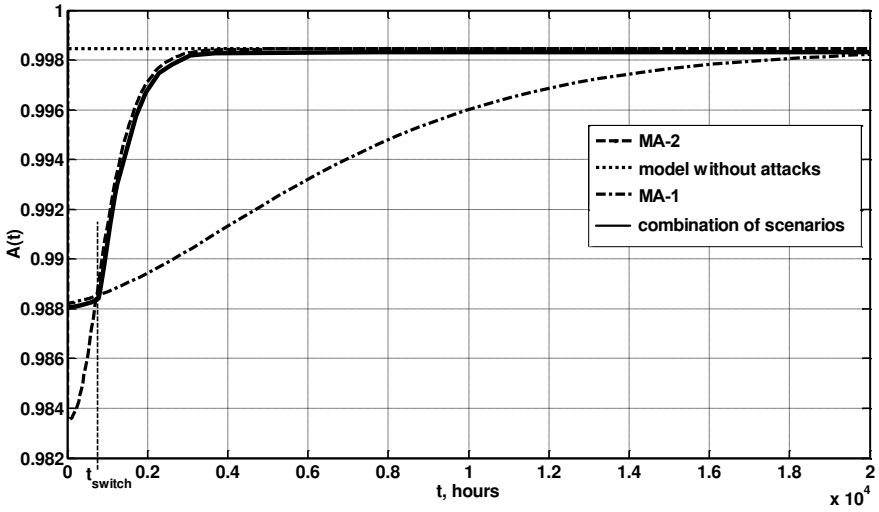


Fig. 9. Combining of the models MA-1 and MA-2 (solid line)

According to Fig. 9, the vulnerability elimination scenario MA-1 is better to use till tswitch =750 hours, after this time the scenario MA-2 ensures better availability. Hence at the beginning recovering a system after attack (without vulnerability elimination) is preferable. Then, taking into account increase of the number of failures caused by attacks other scenario (when vulnerabilities are detected and eliminated both after attacks and during maintenance) becomes preferable. It allows increasing the value of availability from 0.984 (MA-2) to 0.988 (MA-1) and decreasing time transition to stationary state from 20000 (MA-1) to 3000 (MA-2) hours.

5 Conclusions

We analyzed a set of web-system behavior scenarios in conditions of attacks on component vulnerabilities. Quantitative assessment and research of availability for such systems can be based on Markov's models using statistic data about vulnerabilities contained in open databases and described sequence of evaluating of attacks rates and criticality.

We proposed and discussed two models of web-system availability considering attacks on DNS vulnerabilities and different scenarios of vulnerability elimination. There is possibility and reasonability of scenario changing taking into account values of availability function allowing increase minimum one at the non-stationary stage and decrease time of transition to stationary state. This approach allows selecting VE scenario to improve resilience of web-system.

The future research efforts may be concentrated on development of integrated strategies for maintenance and security policies selection taking into account physical, design and interaction faults, and implementation of dynamically reconfigurable web- and cloud-systems with embedded monitor and solver to select the optimal strategy of

maintenance.

Besides, other types of the vulnerabilities for confidentiality and integrity issues and more detailed model taking into account routing processes can be researched.

References

1. Dong Seong Kim, Machida, F., Trivedi, K.S.: Availability Modeling and Analysis of a Virtualized System. In: 15th IEEE Pacific Rim International Symposium on Dependable Computing, pp.365--371, IEEE Press, Shanghai (2009)

2. Zheng Wu, Yang Ou, Yujun Liu: A Taxonomy of Network and Computer Attacks Based on Responses. In: International Conference on Information Technology, Computer Engineering and Management Sciences, pp.26-29, IEEE Press, Nanjing (2011)

3. Roy, A., Dong Seong Kim, Trivedi, K.S.: Cyber security analysis using attack countermeasure trees. In: Proceedings of the Sixth Annual Workshop on Cyber Security and Information Intelligence Research (CSIIRW '10), pp.1--4, ACM, New York (2010)

4. Ping Wang, Jia-Chi Liu: Threat Analysis of Cyber Attacks with Attack Tree+. Journal of Information Hiding and Multimedia Signal Processing 5(4), 778--788 (2014)

5. Alaa Mohammed Abdul-Hadi, Ponochovny, Y., Kharchenko, V.: Development of basic Markov's model research availability of commercial web services. Radioelectronic and computer systems (64), 186-191 (2013)

6. Kharchenko, V., Alaa Mohammed Abdul-Hadi, Boyarchuk, A., Ponochovny, Y.: Web Systems Availability Assessment Considering Attacks on Service Configuration Vulnerabilities. In: Zamojski, W., Mazurkiewicz, J., Sugier, J., Walkowiak, T., Kacprzyk, J. (eds.) Advances in Intelligent Systems and Computing. vol.286, pp. 275--284, Springer International Publishing, Switzerland (2014)

7. Nicol, D., Sanders, W., Trivedi, K.S.: Model-based evaluation: from dependability to security. IEEE Transactions on Dependable and Secure Computing 1(1), 48-65 (2004)

8. Trivedi, K.S., Dong Seong Kim, Roy, A., Medhi, D.: Dependability and security models. In: Proceedings 7th International Workshop on the Design of Reliable Communication Networks (DRCN 2009), pp. 11-20, IEEE Press, Washington, DC (2009)

9. Kizza, J M.: Guide to Computer Network Security. 2nd edition. Springer, London (2013)

10. Al-Kuwaiti, M., Kyriakopoulos, N., Hussein, S.: A comparative analysis of network dependability, fault-tolerence, reliability, security, and survivability. IEEE Communications Surveys & Tutorials 11(2), 106--124 (2009)

11. NVD - Advanced Search, http://web.nvd.nist.gov/view/vuln/search-advanced

12. NVD - Data Feeds, http://nvd.nist.gov/download.cfm#XML

13. Recommendation X.1521. Common vulnerability scoring system. ITU-T, Geneva, The Switzerland (2012)

Author Index